1 Introduction

Towards the end of the 19th century there were two events within the same decade that greatly influenced the way people travel and communicate nowadays. The first one takes place right at the beginning of 1886 in the city of Mannheim in Germany, where Karl Benz applied for a patent of his *Patent-Motorwagen* [Ben86]. The three-wheeled vehicle powered by an internal combustion engine is now widely regarded to be the invention of the modern car. Nowadays, motor vehicles play a fundamental role in modern society from transporting goods to personal travel and recreation. Alone in Germany, as of 2020 there are round 47.7 million registered passenger cars [Kra20]. Current vehicles are however vastly different to the ones produced at the end of the 1800s. Since then, substantial improvements have been made not only in powertrain technology and passenger comfort, but also in the safety systems for both passengers and bystanders. These systems can be generally divided in two categories: passive and active. The former category aims to minimize the severity of a collision and includes e.g. the seatbelt, the airbags and crumple zones. Active systems, on the other hand, are aimed at preventing the collision from happening in the first place. Examples from this category are the anti-lock braking system (ABS), autonomous emergency braking (AEB) and adaptive cruise control (ACC).

The second influential event happened a mere 50 kilometers away from the first one, in the city of Karlsruhe. There, the physicist Heinrich Hertz demonstrated the existence of electromagnetic waves. While their theory had already been laid out by James Clerk Maxwell in 1864, it was not until Hertz conducted a series of experiments between 1885 and 1889 that their existence was proven. This event gave birth to radio communications and its multiple applications, one of them being radar—an acronym for radio detection and ranging. The proto-radar sensor was introduced at the beginning of the 20th century by Christian Hülsmeyer. His *Telemobiloskop* [Chr04], as the sensing device was called, announced the presence of distant metallic objects within its line of sight and its intended use was to avoid the collision of ships. The following developments of radar systems were mainly driven by military necessity, e.g. for surveillance, navigation and weapons guidance [Ric14, pp. 2]. After World War II, civil applications of radar also started to emerge, e.g. weather radar, air traffic control, maritime navigation and by the end of the 20th century in series production cars.

At the beginning, automotive radar systems, such as *Distronic* on the Mercedes S-Class and *Active Cruise Control* on the BMW 7 Series, were used for ACC and were therefore mainly conceived with the comfort of the driver in mind [Gal16, pp. 296]. Later systems, such as AEB, are more concerned with increasing safety, i.e with the prevention of collisions. These kind of safety features are of utmost importance, since there is globally a high number of road traffic casualties. For example, in the EU, where

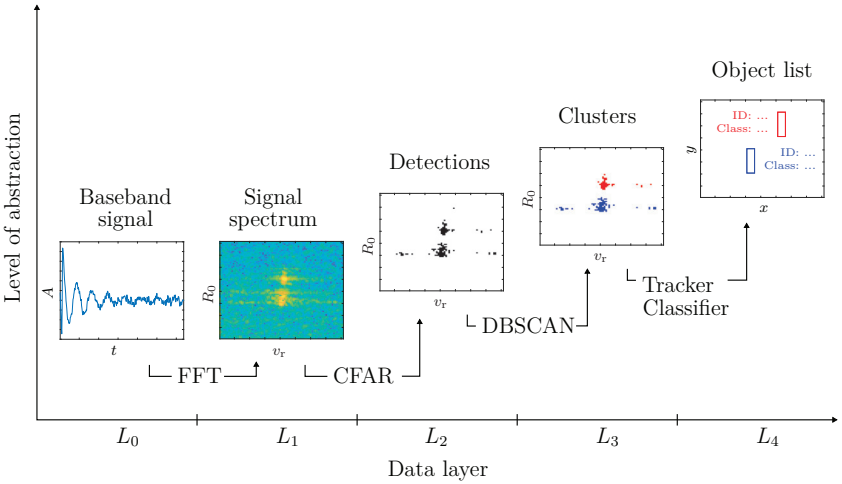

Figure 1.1: Abstraction levels at the different data layers of the classical radar signal processing chain.

the number of road fatalities decreased by around 40% in the span of 10 years, they still summed up to over 25,000 thousand in 2016 [Eur18a]. From this number, 8% is made up from bicycle fatalities [Eur18a] and 21% from pedestrian fatalities [Eur18b]. Pedestrians and cyclists are referred to as vulnerable road users (VRU), since the outcome of an accident is often more severe for them.

From the 21st century onwards, advanced driver-assistance systems (ADAS)—to which the previously mentioned active safety systems are part of—have become increasingly widespread and their functions more sophisticated. From automatic parking systems to blind spot detection and lane departure warning system (LDWS), the list is ever growing. To perform this tasks, radar sensors have become widespread in the automotive industry due to their low cost, high precision measurements of range, velocity and angle, and robustness in adverse weather and lighting conditions. The classical and simplified radar signal processing chain for automotive applications, divided in data layers at the crucial stages, is depicted in Fig.1.1. It all begins with the unprocessed time-domain baseband signal, which contains the information of the targets' range, velocity and angle in its frequency and phase components. By performing a fast Fourier transform (FFT), these frequencies are highlighted and, since the FFT is a linear and invertible function, no information is lost at this stage. A detection algorithm, e.g. constant false alarm rate (CFAR) detection, separates targets from noise and clutter. However, no detection algorithm is perfect, resulting in false positives as well as false negatives, and the higher level of abstraction at layer L_2 comes with information loss. Detections are then gathered in groups (clusters) at layer L_3 using algorithms such as

Table 1.1: The six levels of driving automation, as defined by the SAE International [J3018].

Level	Name	Description
0	No driving automation	All dynamic driving functions are fully performed by the human driver.
1	Driver assistance	Steering or accelerating performed by the driver assistance system under certain traffic situations.
2	Partial driving automation	Both steering and accelerating performed by the driver assistance system under certain traffic situations.
3	Conditional driving automation	The automated driving system performs all aspects of the dynamic driving task under certain traffic situations. The human driver is expected to respond appropriately to a request to intervene.
4	High driving automation	The automated driving system performs all aspects of the dynamic driving task under certain traffic situations, even when the human driver does not respond appropriately to a request to intervene.
5	Full driving automation	The automated driving system performs full-time all aspects of the dynamic driving task under all traffic situations and environmental conditions.

DBSCAN [EKS+96]. Some detections may also be discarded as noise during this stage. Lastly, high level algorithms produce predictions about the targets' attributes, e.g. their physical size, trajectories and classes. This is done based on features extracted at the highest abstraction levels.

For many of the ADAS, the classical radar signal processing suffices. They function mainly in simple and predictable situations, e.g. during highway drives or parking maneuvers. Nonetheless, the end goal in the automotive industry with regards to ADAS is to achieve fully autonomous vehicles. The road to reach this point has been divided into six levels of automation by the SAE International (Table 1.1). From levels 0 to 2, the human driver fully or partially performs the dynamic driving task. At level 2—where most modern vehicles are currently at—the automated driving system (ADS) can control both the steering and accelerating task under certain conditions. With "certain conditions" the previously mentioned simple situations are usually meant. Starting at level 3, the ADS is fully responsible for the dynamic driving task while engaged. Gradually a decreasing readiness of the driver to overtake control is needed until level 5 is reached, where a steering wheel is not even necessary.

With the increasing automation of driving functions comes an increasing demand for the sensors' capabilities. In order for autonomous driving to become a reality, the

vehicle's sensors must be able to function in all kinds of scenarios, not only simple and predictable ones. Especially challenging is the situation presented by urban scenarios. These are more densely populated than motorway scenarios and also contain different types of road users, making them highly dynamic and more complex. For this reason, vehicles will have to accurately capture the environment and also understand it on a semantic level. With the ability to classify the different road users it then becomes possible to apply class specific prediction models, which outperform more general tracking approaches [DLS+19].

The last decade has seen an increase of deep learning techniques being applied to replace established methods across multiple disciplines. For example, in the computer vision domain, the classical approach of performing classification by extracting hand-selected features and then using a shallow classifier have been replaced by deep convolutional neural networks such as AlexNet [KSH12], VGGNet [SZ14] or GoogleNet [SLJ+15]. Also for object detection, approaches that combine the detection and classification in a single step—such as YOLO [RDGF15]—have become prevalent.

The problem of using manually designed features for object detections lies within finding robust ones that fit all kinds of objects with diverse appearance and in different illumination conditions and backgrounds [ZZXW19]. The same can be said about the parametrization of the different algorithms in the classical radar signal processing chain, e.g. of the detection and clustering algorithms, which must fit small and weak targets, such as pedestrians, as well as large and strong targets, such as trucks. Additionally, deep learning algorithms can make use of low-level information, usually discarded by the first steps in the signal processing chain, to perform high-level tasks, such as classification. For this reasons, deep learning has gained considerable interest in the automotive radar signal processing domain. The relevant developments in this area are introduced later in Chapters 4 and 5.

1.1 Goals and Contents of this Work

The goal of this work is to study the feasibility of using deep learning algorithms to enhance, or even replace, steps of the classical automotive radar signal processing chain depicted in Fig. 1.1. To do this, multiple detection and classification approaches, at different data layers of the radar signal processing chain, are developed and implemented.

First, the fundamentals of radar theory are introduced in Chapter 2. This includes the signal model, as well as the classical signal processing chain of continuous wave radar sensors. Additionally, the radar system and its parameters, which are used throughout this work, are presented at the end of the chapter.

Chapter 3 starts with the general theory and terminology of machine learning. After introducing the artificial neuron, deep learning with artificial neural networks and convolutional neural networks is described. Loss and activation functions, as well as common metrics used to evaluate the models in the following chapters are presented.

Two systems for the classification of VRUs are the subject of Chapter 4. A fundamental feature of both systems is the so called micro-Doppler effect, which is described at the

beginning of the chapter. The first system is a novel—at the time of its publication—approach, which uses a convolutional neural network to perform classifications based on single-frame radar measurements. For this, only single-target and relatively simple scenarios are considered. The second system uses a detection procedure to extract regions of interest, which are then classified with a deep learning approach. In this case, the approach is multi-target capable and trained and evaluated with measurement data from test drives in inner city scenarios.

In Chapter 5, two detection systems, one based on time-domain signals (layer L_0) and one based on radar spectra (layer L_1), are presented. The system based on layer L_1 employs a state of the art single-shot object detector, known as YOLO, to simultaneously detect and classify VRUs on 2-dimensional radar spectra. It effectively replaces multiple steps of the radar signal processing chain with a single network: detection, clustering and classification. The second detection system is a novel approach, which runs the time-domain baseband signals through an artificial neural network, which performs detections on the range dimension, effectively replacing the fast Fourier transform (FFT) and the CFAR procedures in the classical signal processing chain.

Lastly, a summary of this work, concluding remarks and an outlook on possible future research are given in Chapter 6.

2 Radar Fundamentals

In this chapter the necessary theoretical background for radar sensors and radar signal processing is introduced. The scope of this work encompasses continuous wave (CW) radars only and for this reason the extent of the theoretical background is also limited to this class of sensors. Hardware aspects are mostly left out in order to focus on the signal models, which build the basis for the algorithms introduced in later chapters.

The derivation of the signal model begins with the simple mono-frequent CW radar, followed by the frequency modulated continuous wave (FMCW) radar and finally the chirp sequence FMCW radar. Following that, target detection procedures on the 2-dimensional radar spectrum are briefly described. These sections are mainly based on Kronauge [Kro14], unless otherwise specified. Furthermore, the signal model for performing angle measurements utilizing a phased array is derived following Richards [Ric14].

The last section in this chapter presents the chosen radar system, the waveform and system parameters selected to achieve the necessary performance in automotive urban scenarios, as well as the hardware modifications needed to integrate it in a test-vehicle.

2.1 Continuous Wave Radar

A radar sensor functions in the broad sense by radiating a high-frequency signal $s_{\mathrm{T}}(t)$ and receiving the backscattered signal $s_{\mathrm{R}}(t)$. The shape of $s_{\mathrm{T}}(t)$, also known as the *waveform*, is determined by the selected modulation scheme. This in turn will determine the characteristics of the radar system in regards to resolution, unambiguous measurement range and multi-target capabilities.

Figure 2.1 depicts the generic block diagram of a CW radar. A waveform generator is used to create a signal, which drives the voltage controlled oscillator (VCO). The VCO then produces the transmit signal, which can be analytically expressed as

$$s_{\mathrm{T}}(t) = a_{\mathrm{T}}(t) \cdot \cos(\phi_{\mathrm{T}}(t)), \tag{2.1}$$

where $a_{\mathrm{T}}(t)$ and $\phi_{\mathrm{T}}(t)$ denote the time dependent amplitude and phase of the signal. In the case of CW radars the amplitude is not modulated, i.e. the time dependent amplitude $a_{\mathrm{T}}(t)$ in Equation 2.1 becomes an arbitrary constant value, which is assumed to be equal to one ($a_{\mathrm{T}}(t) = 1$) to simplify further considerations. The instantaneous transmit frequency $f_{\mathrm{T}}(t)$ of $s_{\mathrm{T}}(t)$ is given by the time derivative of its phase:

$$f_{\mathrm{T}}(t) = \frac{1}{2\pi} \frac{\mathrm{d}\phi_{\mathrm{T}}(t)}{\mathrm{d}t}. \tag{2.2}$$

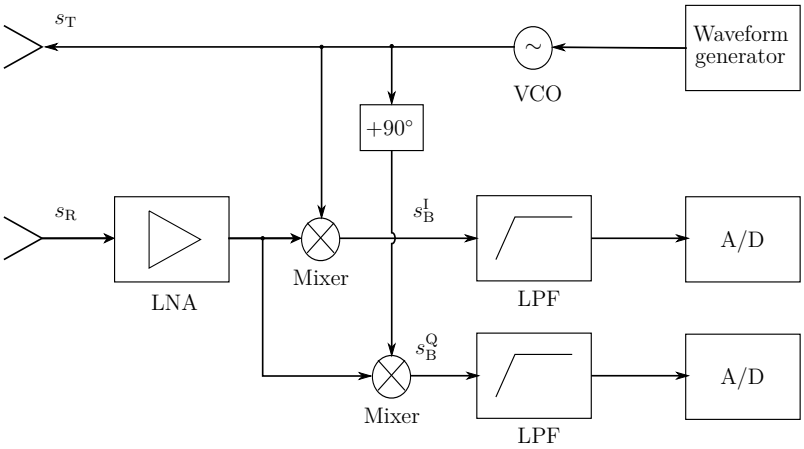

Figure 2.1: Block diagram of a CW radar.

The expression for the transmit signal in Equation 2.1 can then be re-written in terms of $f_T(t)$ using Equation 2.2:

$$s_T(t) = \cos\left(2\pi \cdot \int_0^t f_T(t')\mathrm{d}t'\right). \tag{2.3}$$

If an object is present in the field of view (FOV) of the radar, the receive signal will constitute of the superposition of the backscattered transmit signal at different points on the object. In order to simplify the following derivations, it is assumed that only one point on the object scatters the signal back to the radar. The signal coming from the so called point-target at the radar receiver can thus be regarded as a time-delayed and scaled version of the transmit signal:

$$s_R(t) = A_R \cdot s_T(t - \tau), \tag{2.4}$$

where τ is the round trip delay, i.e. the time that it takes for the signal to travel from the radar to the target and back again to the radar, and A_R is the amplitude of the signal. If the object is located at a range R_0 from the radar, the time delay is

$$\tau = \frac{2 \cdot R_0}{c}, \tag{2.5}$$

with the speed of light $c \approx 3 \cdot 10^8 \mathrm{m\,s^{-1}}$. The amplitude A_R is proportional to the received power P_R, which can be determined using the mono-static radar equation [Sko01, pp. 6]:

$$P_R = \frac{P_T \cdot G_T \cdot G_R \cdot \sigma_r \cdot \lambda^2}{(4\pi)^3 \cdot {R_0}^4}, \tag{2.6}$$

where P_T is the transmit power, G_T the transmit antenna gain, G_R the receive antenna gain, σ_r the object's radar cross section (RCS) and λ the wavelength.

A CW radar simultaneously transmits and receives the backscattered signal. The receive signal is first amplified by a low noise amplifier (LNA) and then down-converted by mixing it with the transmit signal, which produces the intermediate frequency, also known as *baseband*, radar signal $s_\mathrm{B}(t)$. The block diagram in Fig. 2.1 contains an I/Q-mixer, which down-converts the in-phase and quadrature components of the receive signal separately. To simplify matters, the mixers can be regarded as ideal multipliers [Poz05]. In this case, the complex baseband signal can be expressed as

$$\begin{aligned} s_\mathrm{B}(t) &= s_\mathrm{B}^\mathrm{I}(t) - \mathrm{j}s_\mathrm{B}^\mathrm{Q}(t) \\ &= A_\mathrm{R}\Big[\cos(\phi_\mathrm{T}(t)) \cdot \cos(\phi_\mathrm{T}(t-\tau)) - \mathrm{j}\sin(\phi_\mathrm{T}(t)) \cdot \cos(\phi_\mathrm{T}(t-\tau))\Big], \end{aligned} \tag{2.7}$$

where s_B^I and s_B^Q are the in-phase and quadrature components of the complex baseband signal s_B. Using the formula for the product of trigonometric functions [BSMM08, pp. 82] the signal can also be written as

$$\begin{aligned} s_\mathrm{B}(t) = \frac{A_\mathrm{R}}{2}\Big[&\cos\Big(\phi_\mathrm{T}(t) - \phi_\mathrm{T}(t-\tau)\Big) + \cos\Big(\phi_\mathrm{T}(t) + \phi_\mathrm{T}(t+\tau)\Big) \\ &- \mathrm{j}\cdot\sin\Big(\phi_\mathrm{T}(t) - \phi_\mathrm{T}(t-\tau)\Big) + \mathrm{j}\cdot\sin\Big(\phi_\mathrm{T}(t) + \phi_\mathrm{T}(t+\tau)\Big)\Big] \end{aligned} \tag{2.8}$$

The subsequent low-pass filters (LPF) suppress the higher frequency components—those with $\phi_\mathrm{T}(t)+\phi_\mathrm{T}(t+\tau)$ in the argument—of the signal. Therefore, only the difference of the transmit and receive phases remains:

$$s_\mathrm{B}(t) = \frac{A_\mathrm{R}}{2}\Big[\cos\Big(\phi_\mathrm{T}(t) - \phi_\mathrm{T}(t-\tau)\Big) - \mathrm{j}\cdot\sin\Big(\phi_\mathrm{T}(t) - \phi_\mathrm{T}(t-\tau)\Big)\Big]. \tag{2.9}$$

Equation 2.9 can also be expressed using Euler's formula:

$$s_\mathrm{B}(t) = \frac{A_\mathrm{R}}{2}\exp\Big(\mathrm{j}(\phi_\mathrm{T}(t-\tau) - \phi_\mathrm{T}(t))\Big). \tag{2.10}$$

Finally, the signal gets digitalized by the analog-to-digital converters (ADC) so that the subsequent signal processing steps, e.g. detection, clustering or tracking, can be performed.

2.2 Mono-Frequent Continuous Wave Radar

The simplest form of a CW radar is the mono-frequent CW radar. For this kind of waveform, the transmit frequency remains constant at the center frequency f_c for the whole duration of the measurement. Figure 2.2 illustrates the transmit frequency of the mono-frequent CW radar for the duration of one measurement frame T_m.

Plugging $f_\mathrm{T}(t) = f_\mathrm{c}$ into Equation 2.2 yields the transmit signal for the mono-frequent continuous wave waveform (assuming a start phase of zero):

$$s_\mathrm{T}(t) = \cos\Big(2\pi\cdot f_\mathrm{c}\cdot t\Big). \tag{2.11}$$

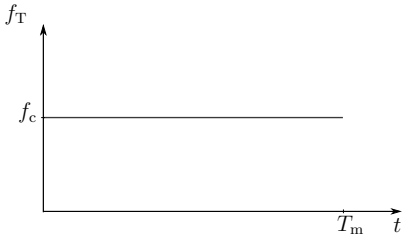

Figure 2.2: Transmit frequency of a mono-frequent CW radar.

Now, the scenario depicted in Fig. 2.3 is used to derive the range and velocity measurements characteristics of this waveform. Again, a point-shaped target assumption is made to simplify the derivations. The baseband radar signal originated by the running person is obtained by substituting $\phi_\mathrm{T}(t) = 2\pi \cdot f_\mathrm{c} \cdot t$ into Equation 2.10:

$$\begin{aligned} s_\mathrm{B}(t) &= \frac{A_\mathrm{R}}{2} \exp\Big(\mathrm{j}2\pi \cdot f_\mathrm{c} \cdot t - \mathrm{j}2\pi \cdot f_\mathrm{c} \cdot \tau(t) - \mathrm{j}2\pi \cdot f_\mathrm{c} \cdot t\Big) \\ &= \frac{A_\mathrm{R}}{2} \exp\Big(-\mathrm{j}2\pi \cdot f_\mathrm{c} \cdot \tau(t)\Big). \end{aligned} \tag{2.12}$$

Since the person is moving relative to the sensor with a radial velocity v_r, the distance becomes time-dependent and thus the round-trip time $\tau(t)$ does too:

$$\tau(t) = \frac{2}{c} \cdot \Big(R_0 + v_\mathrm{r} \cdot t\Big). \tag{2.13}$$

The radial velocity's sign is negative when the target is approaching the sensor (decreasing range and thus round-trip time) and positive when moving away from the sensor (increasing range and thus round-trip time). If the person's movement direction is not radial to the sensor, the radial component of their velocity vector $\boldsymbol{v}$ is given by

$$v_\mathrm{r} = |\boldsymbol{v}| \cdot \cos(\alpha), \tag{2.14}$$

where α is the angle between the heading direction of the person and the boresight direction of the radar.

Using Equations 2.12 and 2.13, the baseband signal can be expressed in terms of the radial velocity and the distance of the target:

$$\begin{aligned} s_\mathrm{B}(t) &= \frac{A_\mathrm{R}}{2} \exp\left(\mathrm{j}2\pi \cdot \left(-f_\mathrm{c} \cdot \frac{2v_\mathrm{r}}{c} \cdot t - f_\mathrm{c} \cdot \frac{2R_0}{c}\right)\right) \\ &= \frac{A_\mathrm{R}}{2} \exp\Big(\mathrm{j}2\pi \cdot \Big(f_\mathrm{D} \cdot t + \phi_0\Big)\Big), \end{aligned} \tag{2.15}$$

where

$$f_\mathrm{D} = -\frac{2v_\mathrm{r}}{c} \cdot f_\mathrm{c} \tag{2.16}$$