2021 International Conference on Control, Automation and Diagnosis (ICCAD) | 978-1-6654-4968-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCAD52417.2021.9638726

Automatic track guidance of industrial trucks using self-learning controllers considering a priori plant knowledge

Timm Sauer, Luca Spielmann, Manuel Gorks and Klaus Zindler
University of Applied Sciences
Aschaffenburg, Germany

Ulrich Jumar
ifak - Institute of Automation and Communication
Magdeburg, Germany

***Abstract*—This paper presents a new self-learning control scheme for lateral track guidance of industrial trucks using artificial intelligence (AI). It is an universally applicable lateral dynamic control concept which is able to adapt itself to different truck variants. Moreover it shall consider vehicle parameter variations that occur during operation, such as the load dependent change of vehicle mass and moment of inertia. The proposed approach uses Reinforcement Learning (RL). In order to reduce the training effort, a new concept is realized, taking into account a priori knowledge of vehicle behavior. Its fundamental idea consists of dividing the training process into two steps. In the first step the controller will be pre-trained on basis of a nominal model representing a priori knowledge of lateral dynamic vehicle behavior. Since this model is derived for an industrial truck with average vehicle parameter values, a fine tuning of the control parameters has to be performed in the second step. In this way the controller is adapted to the actual truck variant and the corresponding vehicle parameter values. In order to demonstrate the efficiency of the proposed control scheme, the simulation results given in this paper are compared to the closed loop behavior using standard LQR.**

***Index Terms*—automatic track guidance, reinforcement learning, self-learning control, a priori knowledge**

I. Introduction

A. Motivation

The automation of intralogistic processes is an important key for ensuring the competitiveness of industrial companies in a global market. Therefore, Aschaffenburg University has been cooperating for many years with Linde Material Handling Ltd., one of globally leading suppliers of automation solutions for intralogistics and one of the most important manufacturers of industrial trucks. The joint research project *Cooperative Autonomous Intralogistic Systems* funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy aims to improve the efficiency of intralogistic processes via intelligent networking and automation of industrial trucks. The cooperative behavior of all autonomous fleet members shall result in a significant increase of internal material flow.

B. Problem Description and Requirements

Figure 1 demonstrates the principle of automatic steering control of an industrial truck. First of all, the desired vehicle trajectory (predefined path) is calculated and stored as data set. The record includes the necessary setpoint information for the automated vehicle guidance, such as the Cartesian coordinates and curvature of the trajectory. In order to ensure a precise vehicle guidance, the forklift has to follow this predefined path with a low lateral deviation a. For this purpose it is necessary to measure the vehicle's position and to calculate the lateral deviation to a reference point on the trajectory. This deviation corresponds in the simplest case to the distance between the vehicle's center of gravity (CoG) and the reference point R in figure 1. Using this information, the controller calculates an appropriate control signal for the steering actuator in order to reduce the lateral deviation of the industrial truck with respect to the predefined path.

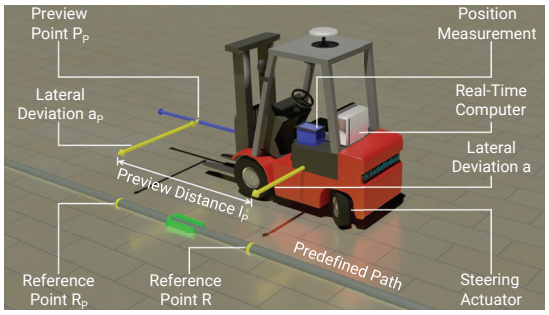


Fig. 1. Principle of automatic track guidance

Control design using classical methods is based on a mathematical model that describes the lateral dynamic vehicle behavior. Hence, in case of the automation of a complete fleet of different industrial truck variants, for each truck variant an appropriate model has to be derived as basis of the control design. This approach is very time-consuming. Moreover, the classical control design methods do not allow the consideration of vehicle parameter changes during operation.
As a consequence, a control concept is required that is capable of controlling the wide variety of different forklift variants and to adapt independently to the actual vehicle parameters.

C. Related Research

In [1] - [5], forklifts are automatically guided along a trajectory using different control methods. However, these publications focus on the automatic track guidance of only one vehicle variant. In addition to the classical adaptive control

concepts ([7] - [10]), new possibilities for adaptive control arise from AI methods. Two different classes of AI control concepts have to be distinguished. The controllers of the first class are called indirect neural controllers [11], [12]. They are characterized by the fact, that an artificial neural network (ANN) is used to model the plant behavior. In a subsequent step this ANN is integrated in the control law. This class includes, for example, the methods of Neural Network Predictive Control [13], [14].
The methods of direct neural control represent the second class of AI control. Here the ANN directly assumes the role of the controller. Starting from an initial parameterization the ANN is optimized using an appropriate training method. Reinforcement Learning (RL) is one of these training methods which especially seems to be well suited to the control problem described above. In analogy to the human learning process RL is performed in closed-loop operation. Through targeted interaction with the vehicle, the controller builds up knowledge of the lateral dynamic vehicle behavior and adapts itself to the actual truck variant and to the current vehicle parameters [15], [16], [17].

However, in view of the large variety of different industrial trucks the RL-control methods known from the literature have an essential disadvantage. Since these methods do not exploit any a priori knowledge about the lateral dynamic vehicle behavior, for each truck variant the whole training process has to be started from scratch. The actually well-known basic vehicle behavior, which is a common feature of all industrial truck variants, must therefore be learned all over again. This results in a unnecessarily time-consuming training process and requires a large amount of training data.

Another disadvantage is related to the training of the controller during real-time vehicle operation. Since RL is started with a random parameterization of the controller without using any a priori knowledge of the basic vehicle behavior, closed-loop stability and hence a safe and collision-free automated vehicle guidance cannot be guaranteed during the training process. Especially in the start-up phase of the training, real-time operation of the RL-controller can be very dangerous.

D. Main Contribution and outline of this paper

In this paper, a novel control concept based on RL for automated lateral dynamic guidance of industrial trucks is presented, which is able to adapt itself to different vehicle variants and vehicle parameters. In contrast to the RL-approaches known from literature, here the existing a priori knowledge of lateral dynamic vehicle behavior will be integrated into the training process. This shall reduce the training effort and improve the robustness of the training process. The main idea consists of dividing the training process into two steps. In the first step the controller will be pre-trained on a nominal model representing a priori knowledge of fundamental lateral dynamic vehicle behavior. Since this model is derived for an industrial truck with average vehicle parameter values, a fine tuning of the control parameters has to be performed in the second step. In this way the controller is adapted to the actual truck variant and the corresponding vehicle parameter values. This step is done in real-time application of the controller. Due to the pre-training phase on the nominal model, a stable closed-loop behavior and hence a safe vehicle guidance is ensured during this second training phase.

Another main contribution concerns the choice of the reference point on the predefined track, used for lateral vehicle guidance. A detailed analysis of the system in section II will show a nonminimum phase behavior, in case of using the vehicle's CoG and the corresponding nearest reference point R on the predefined path for vehicle guidance, see figure 1. This is also an important a priori knowledge, which will be included in the first RL training period using a vehicle model with preview concept that will be explained in detail in section II.

This paper is organized as follows. Section II introduces the control structure. Furthermore, the nominal model used in the first training period of the RL-controller, to include a priori knowledge about the basic lateral dynamic vehicle behavior is derived, validated and analyzed with respect to the influence of the preview concept. In section III the fundamentals of RL are presented and the approach of Twin Delayed Deep Deterministic Policy Gradient (TD3) is explained. Using TD3, an RL-controller is trained in section IV. For comparison purposes additionally a standard LQR is designed. Subsequently, the simulation results of both control concepts are assessed (section V). At the end of the paper, in section VI, the main conclusions and open issues are discussed.

II. Control structure and modeling of the plant

A. Control structure

Figure 2 provides the structure of the proposed vehicle guidance system. The output of the lateral controller δ_{set}, which is the first input signal of the controlled system, is calculated with respect to the reference point R and the vehicle's CoG shown in figure 1, where R is a reference point on the predefined path, which is calculated as a function of the vehicle's CoG by means of the algorithms given in [20]. The curvature $\chi = \frac{1}{\rho_{path}}$ of the predefined path in the reference point R represents the second input of the controlled system, considered as disturbance input of the model. If the preview concept is used, δ_{set} is calculated with respect to R_p and P_p and the path curvature χ_p in R_p represents the disturbance input of the model.

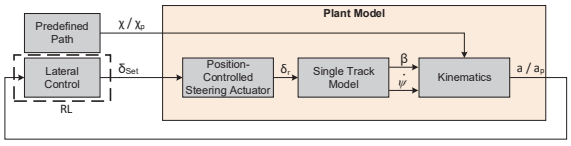


Fig. 2. Structure of the vehicle guidance system

The plant model itself consists of three parts starting with the position controlled steering actuator which gets the calculated setpoints δ_{set} as its input signal and accordingly

adjusts the rear axle steering angle δ_r. The second part of the controlled system is the so-called single track model, that describes the lateral vehicle dynamics depending on the steering angle δ_r. It will be derived in detail in the following subsection II-B. The last part represents the kinematics of the vehicle, i.e. its relative motion with respect to the predefined path. The resulting lateral deviation a, or a_p respectively (output signal of the plant) forms the input signal of the lateral controller.

B. Modeling

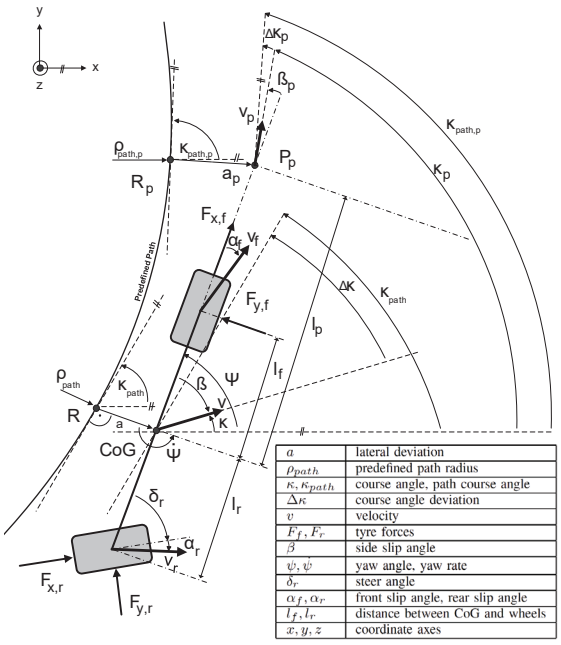
In this subsection the vehicle model with preview as well as the model without preview are derived.
The model of the position controlled steering actuator describes the actuator dynamics in form of a first order delay element with the time constant T_s.

$$\dot{\delta}_r = -\frac{1}{T_s} \cdot \delta_r + \frac{1}{T_s} \cdot \delta_{set} \tag{1}$$

In order to describe the vehicle lateral dynamic, the well-known single track model [18], [20] is adapted to forklifts with rear axle steering. It is based on some simplifications and assumptions:

- Reduction to one wheel per axle
- Neglect of longitudinal dynamic forces such as traction forces, braking forces as well as aerodynamic drag forces
- Constant or only slowly changing vehicle speed
- Small steering angles, slip angles and side slip angles

Taking into account these assumptions and adapting the model to forklifts with rear-axle steering (figure 3), the following equations of motion are obtained.



a	lateral deviation
ρ_{path}	predefined path radius
κ, κ_{path}	course angle, path course angle
$\Delta\kappa$	course angle deviation
v	velocity
F_f, F_r	tyre forces
β	side slip angle
$\psi, \dot{\psi}$	yaw angle, yaw rate
δ_r	steer angle
α_f, α_r	front slip angle, rear slip angle
l_f, l_r	distance between CoG and wheels
x, y, z	coordinate axes

Fig. 3. Single track model with rear axle steering

$$m \cdot v \cdot (-\dot{\beta} + \dot{\psi}) = F_{y,f} + F_{y,r} \tag{2}$$

$$J \cdot \ddot{\psi} = F_{y,f} \cdot l_f - F_{y,r} \cdot l_r \ , \tag{3}$$

where m is the vehicle mass and J is the moment of inertia at the forklifts CoG about the vertical axis. Assuming small steering angles the tire forces $F_{y,f}$ and $F_{y,r}$ can be linearised and represented as:

$$F_{y,f} = c_f \cdot \alpha_f, F_{y,r} = c_r \cdot \alpha_r \ , \tag{4}$$

with

$$\alpha_f = \beta - l_f \cdot \frac{\dot{\psi}}{v}, \alpha_r = l_r \cdot \frac{\dot{\psi}}{v} + \beta - \delta_{set} \tag{5}$$

Thus, the tire forces are assumed to be proportional to the vehicle's slip angles α_f and α_r, while the lateral tire stiffnesses c_f and c_r are assumed to be constant. The third part of the model represents the kinematics of the vehicle and is extending the model equations to describe the relative motion of the vehicle with respect to the predefined path. Specifically, the following relationship results for the lateral deviation and the course angle [19].

$$\Delta\dot{\kappa} = \dot{\kappa}_{path} + \dot{\beta} - \dot{\psi} \tag{6}$$

$$\dot{a} = v \cdot \Delta\kappa \tag{7}$$

$$\dot{\kappa}_{path} = \frac{v}{\rho_{path}} \tag{8}$$

At this point, the influence of the preview concept becomes apparent. By extending the model with the preview concept, the lateral deviation a_p is calculated with respect to the preview point P_p and the reference point on the predefined path R_p (see figure 1) [22]. The equation of the fourth state variable a_p is the only one affected by the extension of the model (see equation (9) which replaces equation (7) in the case of using the preview concept).

$$\dot{a}_p = -l_p \cdot \dot{\psi} + v \cdot \Delta\kappa \tag{9}$$

$$\begin{bmatrix} \dot{\beta} \\ \ddot{\psi} \\ \Delta\dot{\kappa} \\ \dot{a}_p \\ \dot{\delta}_r \end{bmatrix} = \underbrace{\begin{bmatrix} -\frac{c_f+c_r}{m\cdot v} & \frac{-c_r\cdot l_r+c_f\cdot l_f}{m\cdot v^2}+1 & 0 & 0 & \frac{c_r}{m\cdot v} \\ \frac{-c_r\cdot l_r+c_f\cdot l_f}{J} & -\frac{c_r\cdot l_r^2+c_f\cdot l_f^2}{J\cdot v} & 0 & 0 & \frac{c_r\cdot l_r}{J} \\ -\frac{c_f+c_r}{m\cdot v} & \frac{c_f\cdot l_f-c_r\cdot l_r}{m\cdot v^2} & 0 & 0 & \frac{c_r}{m\cdot v} \\ 0 & -l_p & v & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{T_s} \end{bmatrix}}_{\text{system matrix } A} \cdot \begin{bmatrix} \beta \\ \dot{\psi} \\ \Delta\kappa \\ a_p \\ \delta_r \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & v \\ 0 & 0 \\ \frac{1}{T_s} & 0 \end{bmatrix}}_{\text{control matrix } B} \cdot \begin{bmatrix} \delta_{set} \\ \chi \end{bmatrix} \tag{10}$$

The equations (1) - (9) can be given in state space representation for the model with preview concept (10), where $x = \left[\beta, \dot{\psi}, \Delta\kappa, a_p, \delta_r\right]^T$ describes the system's state vector.

u represents the input vector and consists of the steering angle setpoint calculated by the lateral controller and the curvature of the predefined path, considered as disturbance variable.

C. Analysis and validation of the plant model

To investigate the validity of the single track model, the most important part of the model, a double lane change maneuver is performed at two different vehicle speeds. For this purpose, a forklift that is comparable to the nominal Linde E30 (see table I) is equipped with an Inertial Measurement Unit (IMU) and the state variables β and $\dot{\psi}$ as well as the vehicle speed v and the steering angle δ_r are recorded while driving. The measured variables β and $\dot{\psi}$ (solid line) are compared to the corresponding simulation results (dashed line) based on the model derived above (figure 4). The simulation results approximate the time course of the measured variables quite accurately. This a priori knowledge in form of a validated model will be used to pretrain the RL-controller in simulation, in order to build up experience regarding the basic vehicle behavior of a nominal forklift variant. Thus, based on this pre-trained RL-controller, only the fine-tuning has to be done during real time operation, which significantly accelerates the training and reduces the risk of collisions.

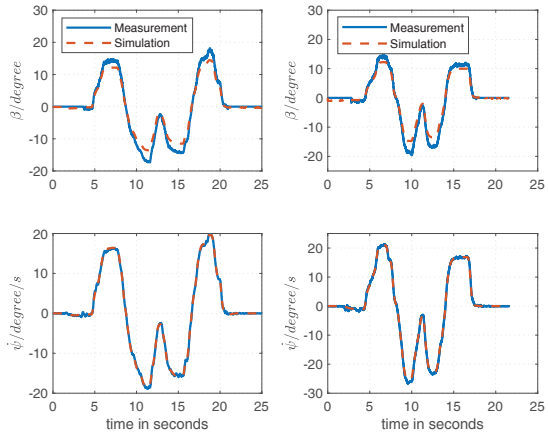


Fig. 4. Double lane change with 1 m/s (left) and 2 m/s (right)

Based on (10) and using the parameter set of a nominal forklift variant like the Linde E30, the pole-zero diagram can be mapped for the system. Figure 5 shows the poles (x) and zeros (o) for the model without preview concept on the left hand side and for the model with preview concept on the right hand side. Obviously, in case of controlling the CoG the diagram shows a zero located in the right half plane. From this, a nonminimum phase behavior can be concluded [1], [10] and results from rear axle steering. Due to this, a more difficult controllability is expected. The effects of the nonminimum phase behavior can be explained by figure 3. While a steering angle, like shown in figure 3 results in a short-time increase in the lateral deviation a at the CoG (reverse response), the lateral deviation a_p at the preview point P_p is immediately reduced by the same steering angle. This effect results from the orientation of the velocity vectors in the vehicle's CoG ($\mathbf{v}$) or the preview point ($\mathbf{v}_p$), respectively. A minimum phase behavior can be ensured by controlling a preview point P_p, located in a sufficiently large preview distance l_p in front of the vehicle [6]. In this case the controller aims to minimize the lateral deviation a_p of the preview point P_p with respect to the reference point R_p, shown in figure 1. Obviously, after integrating the preview point P_p with a constant preview distance of $l_p = 1.5$m, the zero point on the figure 5's right hand side, was shifted to the negative real half-plane. The preview distance is defined to $l_p = 1.5$m in order to place the preview point nearly to the end of the fork.

TABLE I
PARAMETER SETS OF LINDE E16, LINDE E30 AND LINDE E80 [21]

	Linde E16	**Linde E30**	**Linde E80**
m	2984 kg	4981 kg	15720 kg
l	1.492 m	1.665 m	2.400 m
c_f	62000 N/rad	62000 N/rad	62000 N/rad
c_r	122000 N/rad	122000 N/rad	122000 N/rad
l_f	0.705 m	0.858 m	1.181 m
l_r	0.724 m	0.807 m	1.219 m
J	1584 kgm^2	3624 kgm^2	26490 kgm^2
T_s	0.2 sec	0.2 sec	0.2 sec
v	2 m/s	2 m/s	2 m/s

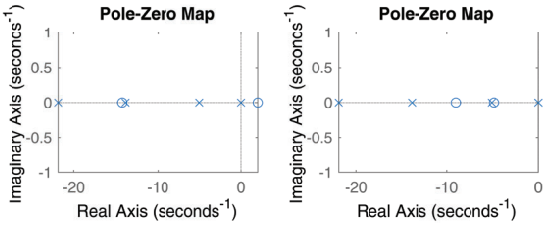


Fig. 5. Pole-Zero-Map of the controlled system without preview concept (left) and with preview concept (right) for Linde E30 forklift

III. REINFORCEMENT LEARNING

A. Basics

Reinforcement Learning in the domain of control systems is not a new approach [23], [24], [25]. Due to the analogy to the human learning process, the self learning characteristics of RL offers potential for solving complex control problems. The principle of closed loop operation process of RL is displayed in figure 6. In RL theory, the basic classical control terms are often replaced. In order to use consistent terms in this publication they are given in table II.

TABLE II
OVERVIEW OF USED TERMS

Classical control theory	**RL theory**
controller	agent
controlled system, plant	environment
control signal	action
state	state, observation

The current state x_k of the vehicle is transmitted to the RL-controller, where comparable to the classical methods, the control signal u_k is calculated in order to affect the controlled system. By this interaction the RL-controller receives the following state x_{k+1} of the vehicle as well as a feedback r_k for the performed output signal in the certain situation, called reward r. Using this information the RL-controller is able to build up knowledge of the lateral dynamic vehicle behavior. This contains information about the relation between the calculated output of the RL-controller on the one hand and the resulting behavior of the controlled system and the corresponding reward on the other hand. Using this information the RL-controller is able to create an internal model of the controlled system during the training process as well as a control policy by itself. The main objective of optimizing the RL-controller's behavior is the maximization of the reward r.

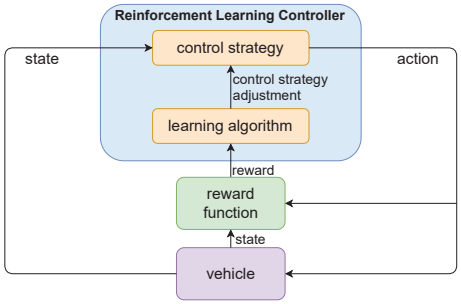


Fig. 6. Principle of Reinforcement Learning

In order to store the knowledge of the RL-controller, different, so-called value functions are used [25]. In this publication $Q_\pi(x,u)$ is called *state action function* and is used to predict the expected reward r of being in the state x and selecting the control signal u. The selected output in a certain system state refers to the strategy π which influences the value function $Q_\pi(x,u)$. In progressive algorithms the strategy is often represented by an ANN. The weights and biases of the ANN act as adjustable parameters θ. The optimal strategy π^* is represented by the optimal ANN parameters θ^* and maximizes the return R (cumulative reward), as simplified in equation (11) [26].

$$\theta^* = \underset{\theta}{\text{argmax}}[R_k] \tag{11}$$

B. Twin Delayed Deep Deterministic Policy Gradient (TD3)

TD3 is a so-called Actor-Critic-method (AC) and will be used in this publication. AC-methods are using separate memory structures to differ between the policy and the value function. Both are represented in form of an ANN. While the actor-ANN is used to calculate the controller's output, the critic-ANN is estimating the value function [25]. This AC-method is known as the TD3 algorithm by [27] and upgrades the Deep Deterministic Policy Gradient (DDPG) algorithm given in [24]. The advantages of the AC, that led to the choice of this method in the context of this publication are briefly listed below [25]:

- Input and output data can be value continuous
- Transition data is stored in a ring buffer which makes the training process more time efficient
- Additional measures, e.g. target-nets are used to stabilize the training process

With TD3, the value function for the critic-ANN is the Q-Function $Q_\pi(x,u)$ and is optimized by methods of supervised learning [28], [29]. Collected state transitions of the controlled system are stored and used for the training of the critic-ANN, in order to map the reward behavior from given state and action as given in equation (12) which includes the discounting of the reward. In applications without a target state, like the continuous task of automatic track guidance of industrial trucks, the discount factor γ ensures a finite value of the sum of the rewards [25].

$$R_k = \sum_{i=k}^{\infty} \gamma^i \cdot r_i \tag{12}$$

The parameters of the actor-ANN should be optimized in order to maximize the state action function $Q_\pi(x,u)$ and thus the expected discounted return R. To implement this, J given in equation (13), is derived based on $Q_\pi(x,u)$. The optimal parametrization of the actor-ANN can be approximated by optimizing J, using a gradient method. θ represents the parameters of the actor-ANN, while ϕ represents those of the critic-ANN [24].

$$\nabla_\theta J \approx \frac{1}{N} \sum_{i}^{N} \nabla_u Q(x,u|\phi)|_{x=x_i,u=\mu(x_i)} \nabla_\theta \mu(x|\theta)|_{x=x_i} \tag{13}$$

IV. Controller Design

In the approach presented in this publication, the main focus is on consideration of existing a priori knowledge. Therefore, the controller design of both methods (LQR and RL-controller) is based on a nominal model with the vehicle parameters of an average industrial truck variant. For this purpose, the parameter set of the Linde E30 is used, which is given in table I. In order to take into account the knowledge regarding the system behavior acquired in section II, the model with preview concept is used in the further course.

A. Classical Approach - Linear Quadratic Regulator

In order to interpret and evaluate the performance of automatic track guidance using the RL-controller, a LQR is additionally designed. Based on the model (see equation (10)) the LQR is designed by the algorithms given in [9], [10]. Here, a quadratic cost function J_{LQR} given in (14), with the diagonal weighting matrices Q_{LQR} and R_{LQR} described in equations (15) and (16) is optimized.

$$J_{LQR} = \int_0^\infty (x^T Q_{LQR} x + u^T R_{LQR} u) dt \tag{14}$$

$$Q_{LQR} = diag([\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5]) \tag{15}$$

$$R_{LQR} = \left[\alpha_6\right] \tag{16}$$

Q_{LQR} weights the quality of the controller while R_{LQR} determines the control energy and thus the controller's dynamic. Both matrices need to be positive semidefinite, which is ensured by diagonal elements $[\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6] \geq 0$ (see table III). The elements of Q_{LQR} are chosen in such a way, that the lateral deviation a_p (fourth state variable) has the highest priority. The lateral deviation is weighted much higher than the other ones by the selection of a suitable factor α_4. Referring to the value α_6 in the R_{LQR} matrix, a high entry has the consequence that the control effort is strongly penalized and the controller becomes increasingly conservative. The minimization of the quadratic cost function results in the control law defined in equation (17), where K is the gain matrix of the LQR.

$$u = -K \cdot x \tag{17}$$

TABLE III
REWARD VARIABLES, RELATED MAXIMUM VALUES AND WEIGHTING FACTORS

variables	max	α
β	1 rad	1
$\dot{\psi}$	2.2 rad/s	1
$\Delta\kappa$	1.57 rad	1
a_p	0.5 m	10000
δ_r	1.53 rad	1
δ_{set}	1.53 rad	5

B. Self-learning Approach - Reinforcement Learning

The LQR was chosen to receive a classical controller, which can be compared to the RL-controller. The two control methods are performing similar, due to the equal cost and reward functions given in (14) and (18). This results in two controllers with identical dynamics [30].
Based on the quadratic cost function of the LQR (14) and the definition of the weighting matrices (15) and (16) the reward function of the RL-controller is given in (18). The sign of this equation indicates a penalty. The goal of the RL-approach is to choose the action of the controller in such a way that the penalty is as small as possible, which is equivalent to maximizing the reward.

$$r_k = -\Big(\alpha_1 \cdot \beta_k^2 + \alpha_2 \cdot \dot{\psi}_k^2 + \alpha_3 \cdot \Delta\kappa_k^2 + \alpha_4 \cdot a_{p,k}^2 + \alpha_5 \cdot \delta_{r,k}^2 + \alpha_6 \cdot \delta_{set,k}^2\Big) \tag{18}$$

The state variables β_k, $\dot{\psi}_k$, $\Delta\kappa_k$, $a_{p,k}$ and $\delta_{r,k}$, used to form the reward function, have to be measured on the real system or on the model in simulation. The last variable used for reward r is the control signal $\delta_{set,k}$ which determines the control effort of the RL-controller. The weighting factors of the reward function and the LQR matrices as well as the variables maximum values are given in table III. In order to take into account the limits of the control signal, a tanh function (19) is used as activation function of the actor-ANN's output layer, which outputs the actuating signal. Due to this, the output is scaled to a range within [-1;1]. By an additional downstream multiplication with the maximum value of the steering angle of 1.53 rad, the limit is defined (compare table III).

$$\tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{19}$$

C. Comparison of the control methods

Both control concepts were designed/pre-trained based on the nominal model using the vehicle parameters of a nominal industrial truck, such as the Linde E30 (table I). In the scenario investigated in the simulation, the vehicle starts with an initial lateral deviation of $a_p = 0.2$m, i.e. offset from the path. The predefined path is applied as a disturbance variable illustrated in figure 7. This means that the path initially runs as a straight line and then changes to a curve with a constant radius. The transition between the mentioned segments is realized using so-called clothoids, where the path curvature is slowly increased until it reaches the final value. The controllers have to compensate the lateral deviation at the beginning and then ensure that the industrial truck precisely follows the predefined path, keeping the lateral deviation correspondingly as low as possible.

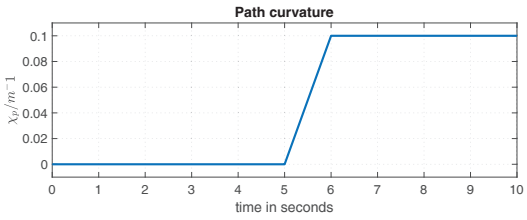


Fig. 7. Path curvature in the reference point R_p

Figure 8 demonstrates the simulation results of both control methods. The lateral offset of the preview point in front of the vehicle a_p (see Fig. 1) as well as the control signals δ_{set} are shown. It can be seen from the figure that both methods have similar dynamic behavior and comparably guide the vehicle.

V. SIMULATION TESTING RESULTS

The previously presented approach of dividing the training process into two steps in order to take into account existing a priori knowledge will now be tested. The control design of the LQR as well as the first training period of the RL-controller takes place in simulation, using the presented vehicle model with preview concept and the parameters of the nominal Linde E30 forklift. In order to investigate the approach with respect to improved training efficiency and improved control quality, the pre-trained RL-controller is used to guide other vehicle variants. For this purpose, both a smaller forklift variant such as the Linde E16 and a larger one, like the Linde E80 are used

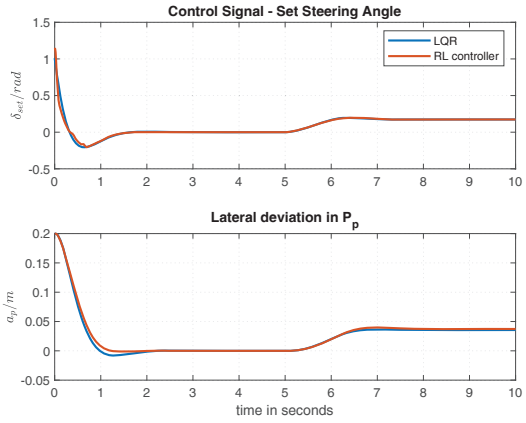


Fig. 8. Comparison of the control concepts

(see table I). Moreover, the adaptability of the RL-controller with respect to changing model parameters is also investigated. For this purpose, the model parameters of the Linde E30 are used and an additional payload of 3 tons is simulated. Instead of a second training period in real-time operation, the simulation model with parameter sets of the corresponding variants are used in this paper. Table IV impressively shows that there are considerable advantages in terms of training efficiency as a result of pre-training the RL-controller based on the nominal model. Since the well-known basic vehicle behavior has already been trained, the RL-controller is able to adapt the control parameters in the second training period to the new vehicle parameters by means of a short downstream training, without the need to carry out the entire experiences all over again. This increase in efficiency is clearly evident in all three studies within this section (see the optimization steps in table IV).

TABLE IV
OVERVIEW TRAINING CONDITIONS

Training	Epochs	Optimization steps
Training done in advance - E30	150	116822
Follow up training - E16	6	6000
Follow up training - E80	14	14000
Follow up training - E30 - 3t payload	2	2000

The performance of the RL control concept will again be compared with the LQR designed for the Linde E30, which is used as a representative of the classical control methods as well as with the pre-trained RL-controller without fine-tuning. For this, the scenario given in figure 7 is used again. Since the vehicle variants of the Linde E30 and the Linde E16 are comparable, both controllers work very well (see figure 9). The retrained RL-controller can adapt the control parameters in a short downstream training using the parameter set of the Linde E16. The advantage becomes particularly clear in the rear course of the diagram, when the path curvature is applied. After the second training period the RL-controller can significantly reduce the steady-state control error. Due to the basic parameterization of the RL-controller, based on the first training period, the second one can efficiently be done within a very short time (6000 optimization steps).

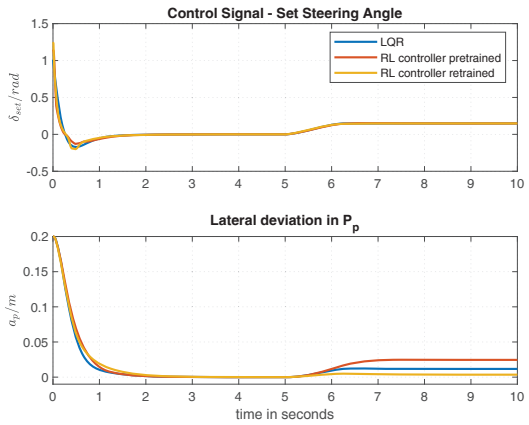


Fig. 9. Truck variant Linde E16

In the next simulation study, the Linde E80 is used. This forklift variant is one of the largest and differs significantly from the E30, which is reflected in the course of the LQR and the pre-trained RL-controller. Both suffer significant control quality losses, see figure 10. Oscillations can be seen in the control signal and thus also in the lateral deviation a_p. In addition, the amplitude also increases significantly. It can be seen that both controllers cannot completely compensate the initial lateral deviation from the path within the first 5 seconds, before the trajectory entered the curve. The self-learning controller shows its advantages particularly clear. Through a short downstream training, it is able to adapt to the new vehicle variant. The self-learning controller guides the vehicle from the initial lateral deviation quicker onto the nominal trajectory while also exhibiting improved damping. The control quality during the curve also shows clear advantages.

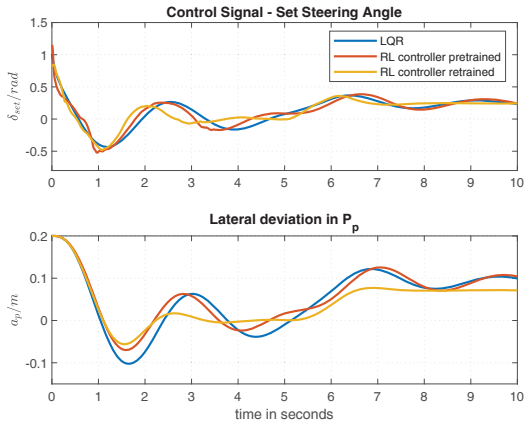


Fig. 10. Truck variant Linde E80

Finally, it is to be tested whether the control systems can also react to varying plant parameters, such as an additional payload. For this purpose, the Linde E30 parameter set is used

and an additional weight of 3 tons is assumed in simulation. The RL-controller can adapt again to the new vehicle variant during the second training period and guides the vehicle from the initial lateral deviation without overshooting onto the nominal trajectory. The control quality during the curve also shows clear advantages compared to the other two controllers (figure 11).

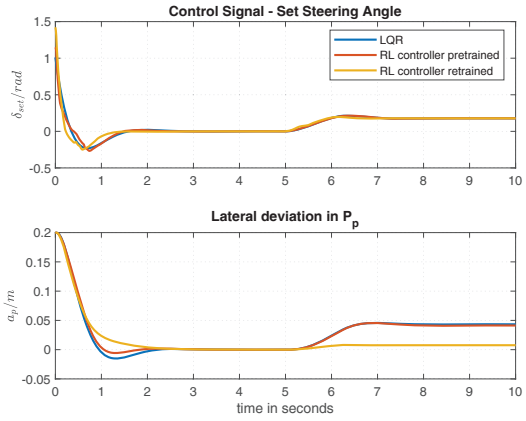


Fig. 11. Truck variant Linde E30 with additional load of 3 tons

VI. Conclusion and Future Work

This paper proposes a new self-learning control method for an accurate track guidance of industrial trucks. This approach is based on RL-methods, taking into account the existing plant knowledge during the training process. The advantage of the well-known plant behavior is based on basic physical laws and the relationships derived in the modeling. It is integrated into the training process in the form of an experience build-up of the self-learning controller carried out in the simulation. The adaptation to new industrial truck variants is realized in this paper with downstreamed trainings in simulation, using the same single-track model with different vehicle variant parameter sets. This means that the control parameters only have to be slightly adjusted, which ensures a certain degree of safety and shortens the training process enormously. Moreover, a LQR based on the linear single track model was designed in order to assess the performance of the developed RL control scheme. The performance of the different controllers were compared in simulation tests and demonstrate the ability of the RL-controller to adapt to different vehicle variants and vehicle paramters. Due to the basic parameterization, during the first training perdiod, the RL-conroller is able to adapt itself to different truck variants and is able to consider vehicle parameter variations.

References

[1] Longqing Li, Kang Song, Hui Xie: Path-Following Control for Self-driving Forklifts based on Cascade Disturbance Rejection with Coordinates Reconstruction, Proceedings of th 39th Chinese Control Conference, 2020.

[2] Tua Agustinus Tamba, Quyen T. T. Bui, Keum-Shik Hong: Trajectory Generation of an Unmanned Forklift for Autonomous Operation in Material Handling System, SICE Annual Conference, 2008.

[3] Ritzer et al.: Advanced Path Following Control of an Overactuated Robotic Vehicle, IEEE Intelligent Vehicle Symposium (IV), 2015.

[4] Muhammad Raheek Bhutta, Keum-Shik Hong: Collision-Free Navigation of Forklifts by Points-of-Interest Switching, International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 2012.

[5] Mohammadi et al.: Model Predictive Motion Control of Autonomous Forklift Vehicles with Dynamics Balance Constraint, 14th International Conference on Control, Automation, Robotics and Vision, 2016.

[6] Tan et al.: Developement of an Automated Steering Vehicle Based on Roadway Magnets - A case Study of Mechatronic System Design, IEEE/ASME TRANSACTIONS ON MECHATRONICS, Vol. 4, 1999.

[7] I. Landau, R. Lozano, M. M'Saad and A. Karimi: *Adaptive Control: Algorithms, Analysis and Applications*, Springer, London, 2013.

[8] K.J.Aström, B. Wittenmark: *Adaptive control*, Mass: Addison-Wesley, Reading, 2. Edition, 1995.

[9] H. Unbehauen: *Regelungstechnik III. Identifikation, Adaption, Optimierung*, Vieweg+Teubner Verlag, Wiesbaden, 7. Auflage, 2011.

[10] W. Levine and T. Sawa: *The Control Handbook*, CRC PRESS, IEEE Press, 1996.

[11] A. Levin, K. Narendra: *Control of Nonlinear Dynamical Systems Using Neural Networks: Controllability and Stabilization*, IEEE Transactions on Neural Networks, Vol. 4, No. 2, 1993.

[12] W. Choromanski: *Application of Neural Network for Intelligent Wheelset and Railway Vehicle Suspension Designs*, Vehicle Systems Dynamics, 25:S1, 87-98, 1996.

[13] F. M'Sahli, R. Matlaya: *A neural network model based predictive control approach: application to a semi-batch reactor*, Int. J. Adv. Manuf. Technol. (2005) 26: 161 - 168. Springer London, 2005.

[14] L. Cheng, W. Liu, Z. Hou, J. Yu and M. Tan: *Neural-Network-Based Nonlinear Model Predictive Control for Piezoelectric Actuators*, IEEE Transactions on Industrial Electronics, Vol. 62, No. 12, 2015.

[15] A. Sallab, M. Abdou, E. Perot, S. Yogamani: *End-to-End Deep Reinforcement Learning for Lane Keeping Assist*, 30th Conference on Neural Information Processing Systems (NIPS), 2016.

[16] I. Kageyama, Y. Owada: *An Analysis of a Riding Control Algorithm for Two Wheeled Vehicles with Neural Network Modelling*, Vehicle System Dynamics, 25:S1, 317-326, 1996.

[17] I. Yoshiaki, S. Toshiyuki: *Neural Network Application for Direct Feedback Controllers*, IEEE Transactions on Neural Networks, Vol. 3, No. 2, 1992.

[18] H. Pacejka: Tyre and vehicle dynamics (2nd ed.), Oxford: Butterworth-Heinemann, 2006.

[19] I. Söhnitz: Querregelung eines autonomen Strassenfahrzeugs. Fortschr.-Ber. VDI Reihe 8, Nr. 882. VDI Verlag Düsseldorf, 2001.

[20] Zindler et al.: Querdynamische Fahrzeugführung zur reproduzierbaren Erprobung von Sicherheitssystemen. at-Automatisierungstechnik 60(2), Oldenbourg-Verlag, S. 61-73, 2012.

[21] LINDE. Homepage Linde Material Handling. Accessed on 22.06.2021.

[22] L. König et al.: Nichtlineare Lenkregler für den querdynamischen Grenzbereich (Nonlinear Steering Controllers for the Lateral Dynamics Stability Limit). In: Automatisierungstechnik 55 (2007), Nr. 6.

[23] M. Vogt: *An overview of deep learning techniques*, at - Automatisierungstechnik 66 (9), pages 690-703, Oldenbourg Wissenschaftsverlag, 2018.

[24] T. Lillicrap et al.: Continuous Control with Deep Reinforcement Learning, International Conference on Learning Representations, 2016.

[25] R. Sutton and A. Barto: Reinforcement Learning: an introduction. The MIT Press, 2018.

[26] S. Havenstrom et al. "Proportional integral derivative controller assisted reinforcement learning for path following by autonomous underwater vehicles." arXiv preprint arXiv:2002.01022, 2020

[27] S. Fujimoto, H. van Hoof and D. Meger: Addressing Function Approximation Error in Actor-Critic Methods, International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

[28] K. Gurney: An introduction to neural networks, UCL Press Limited, 1997.

[29] M. Hagan, H. Demuth, M. Beale and O. De Jesus: Neural Network Design, Martin Hagan, 2014.

[30] Y. Ichikawa and T. Sawa: *Neural Network Application for Direct Feedback Controllers*, IEEE Transactions on neural networks, VOL. 3, NO. 2, 1992.