INTERNATIONAL DESIGN CONFERENCE – DESIGN 2024
https://doi.org/10.1017/pds.2024.216

Minimizing occupant loads in vehicle crashes through reinforcement learning-based restraint system design: assessing performance and transferability

Janis Mathieu [1,2,✉], Parul Gupta [3], Michael Di Roberto [1] and Michael Vielhaber [2]

[1] Porsche Engineering Group GmbH, Germany, [2] Saarland University, Germany,
[3] Ilmenau University of Technology, Germany

✉ janis.mathieu@porsche-engineering.de

Abstract

The optimization of mechanical behavior in safety systems during crash scenarios consistently poses challenges in vehicle development. Hence, a reinforcement learning-based approach for optimizing restraint systems in frontal impacts is proposed. The trained agent, which adjusts five parameters simultaneously, is capable of minimizing loads on a seen and unseen anthropomorphic test device on the co-driver position and is thus able of transferring knowledge. A hundred times higher rate of convergence to reach a similar optimum compared to a global optimization algorithm has been achieved.

Keywords: data-driven design, computational design methods, occupant safety, optimization

1. Introduction

With more than one million people killed on the roads worldwide in each year over the past decades, the consideration of safety-relevant aspects to better protect passengers and pedestrians will continue to play a crucial role in the development process of new vehicle generations (Gonter et al., 2021). To improve the crashworthiness of vehicles, structures such as side sills, longitudinal and cross members, but also the restraint system, which is particularly relevant for occupant protection, are optimized regarding the best possible mitigation of injury risks. The considerations to address the optimization of mechanical system behavior are carried out virtually with the aid of numerical finite element (FE) simulations since physical tests of individual components or entire prototypes are highly time- and resource-intensive (Schuhmacher, 2020; Gonter et al., 2021). This enables concepts to be evaluated more quickly which results in more agile and efficient development processes. Additionally, physical crash tests serve for the validation of virtual models.

During the vehicle development process, multiple objectives from various domains must be considered and addressed. For passive vehicle safety, this necessitates conducting numerous investigations to disentangle complex relationships and dependencies. The resulting large number of required simulations, coupled with the intricacy and size of FE models, also render virtual assessments computationally- and thus cost-intensive (Schuhmacher, 2020; Gonter et al., 2021). To further digitalize and increase the efficiency of modern development processes, recent focus has shifted towards data-driven approaches and concepts (Langen et al., 2022). Predictive models can assist in the very early phases by providing more precise insights for subsequent product generations based on prior product generations as discussed for the assessment of crash pulses in the context of passive vehicle safety (Rabus et al., 2022). Large crash simulation datasets generated through the multitude of analyses can be evaluated more expeditiously through intelligent algorithms (Kracker et al., 2023). Subsequently, these approaches hold significant potential within the field of vehicle safety, but also across other domains of vehicle development such as the calibration of thermal systems (Muhl et al., 2022) or engine control units (Rudolf et al., 2022).

The proposed methodology within this work centers on the aspect of automated decision-making in conjunction with the resolution of optimization problems aimed at minimizing occupant loads during a crash scenario. Following a review of the current state of the art, research gaps are identified (Section 2) and subsequently addressed through the presented approach (Section 3 and 4).

2. Design optimization of mechanical systems

Optimizing the behavior of crash loaded systems in passive vehicle safety is particularly important due to the relevance for legal regulations and, at the same time, exceedingly complex attributable to large deformations, nonlinear material and contact behavior. Subsequently, structural responses are often highly nonlinear and sensitive (Schuhmacher, 2020). During the development process, simultaneous optimization of the vehicle structure and the restraint system is performed with the aim of ensuring the integrity of the structure during a crash and reducing the loads on anthropomorphic test devices (ATD) (Schuhmacher, 2020; Gonter et al., 2021).

For optimizing mechanical system behavior global optimization algorithms (e.g., genetic algorithms), local optimization algorithms (e.g., gradient methods), but also heuristic-based approaches can be applied (Schuhmacher, 2020). Especially for shape and topology optimization of structural members like the cross section of a side sill graph-heuristic approaches as discussed in detail by Beyer et al. (2020) have shown significant potential. Objectives for structural optimization can be maximizing the energy absorption or the reduction of acceleration peaks during the crash load.

The restraint system, especially airbags, and seatbelts, are in most cases described with certain parameters in the virtual model such as a restraint force or the outlet valve. The seating position also has influence on the loads and is especially interesting for autonomous vehicles (Huang et al., 2015). Approaches discussed in literature for optimizing restraint system parameters include the response surface methodology (Thiele et al., 2006) or the use of a global optimization algorithm either with a surrogate model (Horii, 2017; Joodaki et al., 2021) or a computationally efficient simulation model (Huang et al., 2015).

A literature overview on the optimization of mechanical systems using reinforcement learning is presented in Section 2.1 and considered to be particularly relevant for the presented work. In the subsequent Section 2.2, the design of restraint systems is discussed. Based on this, the objectives for the design of reinforcement learning (RL)-based restraint system optimization are specified in Section 2.3.

2.1. Optimization of mechanical systems with reinforcement learning

The primary objective of an RL algorithm is to learn an optimal policy for an RL agent that maximizes the cumulative reward over time during interaction with an environment. Through a process of exploration and exploitation, the agent learns which actions lead to favorable outcomes and adjusts its policy accordingly. According to Equation 1, the policy π corresponds to the conditional probability distribution $P(a|s)$ which action a to take, when the environment state is s. The RL agent uses feedback in the form of rewards to adapt its behavior, aiming at improving its decision-making abilities and achieving long-term goals (Sutton and Barto, 2018).

$$\pi: A \times S := \{(a,s)|a \in A, s \in S\} \rightarrow [0,1]$$
$$\pi(a,s) = P(a|s) \tag{1}$$

Hayashi and Ohsaki (2020) analyzed the topology optimization of truss structures under static stress and displacement constraints with the objective of minimizing the structural volume. In an automated process, the RL agent interacted with a FE simulation pipeline to evaluate the structural behavior. Graph embeddings were used to provide a uniform representation of the structures, since the used convolution operations cannot be applied to discrete structures due to their irregular connectivity. The agent showed generalization capabilities on truss structures with different sizes and unforeseen boundary conditions. Hence, the RL agent was able to learn its own rules for optimization in the complex task. The

https://doi.org/10.1017/pds.2024.216 Published online by Cambridge University Press

computational costs of the trained agent were significantly lower compared to a genetic algorithm that has been used for benchmarking. Notable is that the training is computationally expensive, but only needs to be done once. In further research, Hayashi and Ohsaki (2022) applied the technique to the optimization of planar steel frames under short- and long-term elastic load conditions, whereby the agent aims to minimize the structural volume under several practical constraints. In each action, the agent specifies each member by choosing it from a prescribed list. The trained agent outperforms a particle swarm optimizer in time and design quality for cross-sectional design changes. Laborious training and hyperparameter tuning processes have been identified as significant limitations of the approach. A single training process of the agent took tens of hours for both problems due to the sequential execution of FE simulations and subsequent evaluation of the performance of the structure.

Trilling et al. (2022) analyzed the local structural stiffness optimization of graph structures with the training of an RL agent. The actions contain changes to a structure formulated in a uniform graph representation. The observation contains pictures of the deformed model, and a stiffness metric is used to calculate the reward. The agent showed great potential in stiffening four different structures. However, a study for evaluating generalization capabilities was not performed.

Brown et al. (2022) used an RL approach to optimize the topology by removing elements of a discretized planar block structure. The agent observes three channels describing the von mises stress, element bounding, and loading of the structure with a convolutional neural network. The reward is calculated based on the strain energy and the voided elements. The generalization capabilities were tested by optimizing different structure sizes. A benchmark compared to a gradient-based optimizer showed that in five out of six load cases better results in terms of volume fractions could be archived. A limitation is the computational effort of the RL method due to the high amount of FE simulation executions of up to 144 needed to determine the strains under a load in each episode.

The current research in the field of mechanical system optimization with RL shows significant potential, particularly in the transfer of a pre-trained policy to a similar problem. Existing data can be used to reduce the simulations needed to find an optimum. If the FE models need a vast computational resource due to their size and complexity costs can also be sizably reduced. In some cases, the use of RL might lead to sub-optimal solutions (Brown et al., 2022), but in contrast for one structure in the work of Hayashi and Ohsaki (2020) the RL agent was able to find a better solution than the benchmark algorithm. However, in practical applications, sub-optimal solutions, or a direction of where an optimal solution is located might be sufficient and can be used for further analysis with another algorithm or by the engineers themselves.

2.2. Optimization of restraint systems

The ATD loads to be minimized by adjusting the restraint system configuration are typically part in either statutory regulations such as FMVSS 208 or consumer protection guidelines such as the Euro-NCAP (Gonter et al., 2021). Relevant values are, for instance, the head injury criterion in an interval of 15 ms (HIC15), neck injury criterion (Nij) or the maximal chest acceleration. The restraint system needs to be optimized such that load limits given by the regulations are fulfilled in various load cases. One of the dominant load cases for the restraint system design is the frontal impact due to the complex interaction of seatbelt and multiple airbags as well as the unfavorable vehicle kinematics for the ATD. Typically, the airbag and seatbelt models are parameterized allowing a fast adaption of the simulation input files as well as a coupling to commercial optimization software (Thiele et al., 2006; Joodaki et al., 2021).

Xue et al. (2014) compared different optimization algorithms and their application to restraint system design. Simplified rigid body models were used, which are rarely used in modern vehicle development processes. Genetic algorithms were found to propose the most optimal solutions in the least number of steps, due to the presence of a large number of ordered and unordered discrete variables. Gradient-based methods are not able to find satisfying solutions, due to the local properties. Simulated annealing takes the longest time to converge, resulting in a high number of simulations of almost 4000. The best investigated genetic algorithm needed around 2000 simulations for convergence, which is also conditionally feasible especially when the simulation models are larger and more detailed.

Due to this reason a conventional approach of optimizing restraint system parameters is the successive response surface methodology (Thiele et al., 2006). Here, the iterative use of response surface models and the reduction of the area around the optimum defined in the previous step result in a more efficient way of determining an optimum. The method has better exploration than a purely gradient-based method but still might not reach a global optimum.

Newer approaches rely on global optimization algorithms in combination with surrogate models, such as the Gaussian Process model used by Horii (2017). The execution of the machine learning model in less than a second is significantly faster than a FE simulation. However, the validity of these models is only given in the specific problem statement they have been trained for. Horii (2017) was able to specify the pareto optimal solution for HIC15 and maximal chest acceleration with the use of an evolutionary algorithm. As discussed by Thiele et al. (2006) and Joodaki et al. (2021) surrogate models can also be used for design exploration and manually specifying solutions. Joodaki et al. (2021) tested a variety of models including lasso regression and neural networks, whereby the models were trained with 450 large-scale FE occupant LS-DYNA sled models. An ensemble machine model was able to deliver the highest accuracy. The lasso regression model showed reasonable accuracy, while the hyperparameters are easy to adjust.

Rabus et al. (2022) presented an approach for predicting a surrogate value ROLCp, which shows a significant correlation with the chest acceleration, based on a vehicle crash pulse and vehicle configuration parameters. Hence, the model can help to evaluate crash pulses in a very early phase and can thereby contribute to steering structural development in an optimal direction from an occupant safety perspective without the presence of occupant safety simulations.

2.3. Derivation of research objectives

The present work focusses on the mechanical design phase of restraint systems in vehicle development as it is discussed in the work of Thiele et al. (2006), Xue et al. (2014), Huang et al. (2015), Horii (2017) and Joodaki et al. (2021). Limited generalization capabilities from one problem statement to another can be identified as a major shortcoming in the reviewed approaches. Especially since automotive companies typically develop a variety of vehicle series and derivatives. Subsequently, design tasks, as well as optimization problems are recurrent and therefore arise in different models and generations. As a result, a vast amount of simulation data is generated which serves no further purpose after evaluation. For optimizing similar problem statements, RL has shown substantial potential in structural optimization of two-dimensional topologies (Brown et al., 2022), academic extrusion profiles (Trilling et al., 2022) or binary truss topologies (Hayashi and Ohsaki, 2020). Hence, the present work proposes a methodology for the RL-based restraint system design aiming at minimizing loads on ATDs in a frontal impact crash scenario and subsequently tries to provide proper answers for the following three research questions:

1. Can the RL agent effectively acquire the capability to minimize ATD loads through design changes in the restraint system?
2. In how far can the learned policy be transferred and utilized for similar problem statements, such as the use of a different ATD type?
3. How can the performance evaluation of the RL agent be conducted using conventional optimization approaches, particularly regarding the handling of large-scale industry-standard FE simulation models?

The proposed approach in the data-driven design paradigm is aimed at the reduction of the number of simulations needed for restraint system optimization during vehicle development and stores knowledge in terms of recommended actions in the algorithm.

3. Reinforcement learning-based restraint system design

This section introduces the problem statement and the use case along which the training and evaluation of the RL agent is explained. A detailed overview of the RL agent setup and performance is given. The obtained results in regard to the formulated research questions in the previous section are discussed in the following section.

https://doi.org/10.1017/pds.2024.216 Published online by Cambridge University Press

3.1. Problem statement and use case

For optimizing restraint systems in frontal impact crash scenarios sled models as in the work of Thiele et al. (2006) and Joodaki et al. (2021) are used. Due to the application of a predefined vehicle movement onto the body in white and the reduced geometry these models are less computationally expensive than a full vehicle simulation but are still significantly more detailed than simplified MADYMO models as used by Huang et al. (2015) or Horii (2017). The analyzed use case for this work is a frontal impact crash test according to FMVSS 208 with two different ATDs (Hybrid III 50% male ATD (H3) and Hybrid III 5% female ATD (HF)) on the co-driver seating position. To reduce runtime and use pre-existing benchmark simulation models as a foundation, a prediction model based on a design of experiment is trained to describe functional relations between restraint system parameters and ATD loads. This allows for a quick training and evaluation of the RL agent within two problem statements due to the less computational effort in each training or evaluation step. Compared to one simulation model which requires between 10-20 hours on a high-performance computing cluster, the ATD loads for two different ATDs can be specified in real-time with the trained prediction model. As visualized in Figure 1, a reduced set of restraint system parameters is assumed to be relevant due to significant influence on the ATD load (Thiele et al., 2006; Joodaki et al., 2021). In the airbag (AB), the vent hole size as well as the time for the vent hole size adaptivity is adjusted. For the seatbelt (SB), the two force levels as well as the switching time from one to another level are part of the investigation. The head injury criterion (HIC15), the maximum chest acceleration over an interval of 3 ms (a3ms), and the residual distance between the head and dashboard (RA) have been identified as relevant parameters for evaluating the ATD load. For the optimization problem, it is particularly interesting since the output parameters are counteracting as the RA needs to be increased while HIC15 and a3ms need to be reduced.

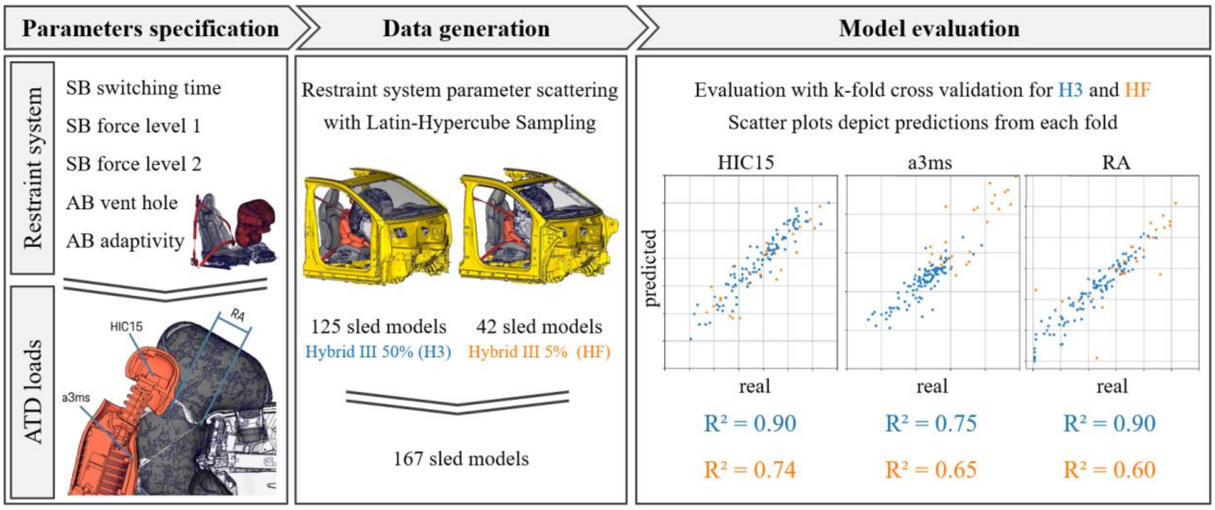

Figure 1. Restraint system parameters, data generation, and prediction model metrics

In total, 167 sled models are used for the training of the gradient tree boosted model. The performance is evaluated with a k-fold cross validation, whereby the coefficient of determination (R^2) is calculated at the end to include the predicted values for each datapoint predicted in one of the folds. This allows for a performance measurement in the entire data space. The scatter plots visualizing the predictions from all the folds for HIC15, a3ms and RA is given in Figure 1. For the H3, the results with an R^2 of up to 0.90 are better than the R^2 for the HF of up to 0.74. This behavior of the model is attributable to the smaller number of datapoints available for the HF. The performance of the model for both ATDs is sufficient since the model mainly serves for the verification and validation of the RL methodology.

3.2. Reinforcement Learning-based methodology

In the first step, the optimization problem described in Figure 1 is converted to a RL task within a Markov Decision Process, whereby states, actions, and rewards are defined as visualized in Figure 2. The action contains a set of changes to the parameters. The observation space is defined as the system

https://doi.org/10.1017/pds.2024.216 Published online by Cambridge University Press

model input and output parameters and the reward is defined based on relative deviations. Since the actions specify relative changes, the agent is trained to adapt the ones initially specified, since the values might differ between vehicle series or generations.

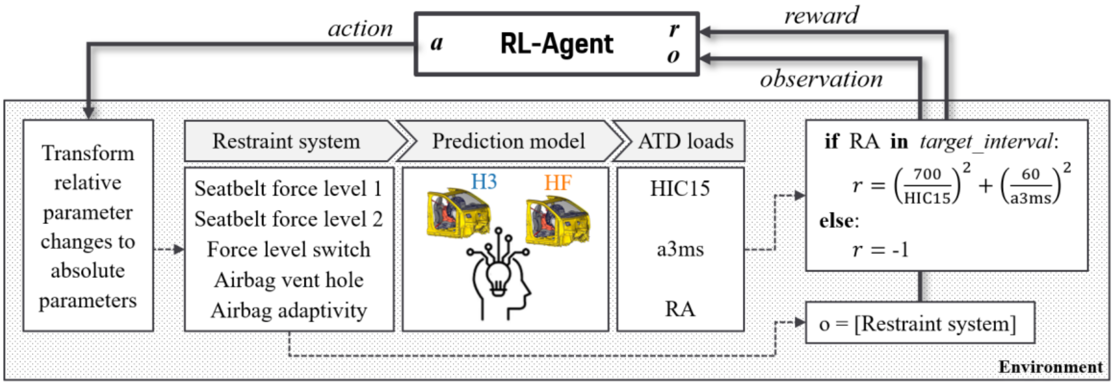

Figure 2. RL agent and custom environment

3.2.1. Reinforcement learning agent

For the RL agent a policy proximal optimization (PPO) algorithm is used. The ability to strike a balance between exploration and exploitation is suitable for the task of discovering policies. It incorporates inherent stochasticity into its policy, which encourages exploration while maintaining the convergence properties needed to find optimal policies. In numerous applications PPO has been identified as highly sample efficient (Beysolow II, 2019). Additionally, PPO can handle discrete action spaces, making it particularly well-suited for the relative changes which are learnt during training. This emulates the decision-making behavior of engineers, who might, for instance, state that the first seatbelt force limit needs to be reduced by 0.25 kN.

3.2.2. Custom environment

In most cases RL, relies on pre-existing environments that offer standardized tasks for agents in common scenarios (Sutton and Barto, 2018). However, custom environments become essential when these pre-existing ones do not fit the problem at hand. For the proposed framework a custom environment to simulate crash scenarios, providing a detailed evaluation of the RL agent's performance is created. Custom environments offer flexibility in defining state space, action space, rewards, rules, and dynamics (Beysolow II, 2019). In this context, the **action space** exhibits five dimensions, each offering three discrete actions for modifying the restraint system configuration. Variables can either be increased, lowered, or be kept constant, resulting in three possible actions for each of the variables. The **observation space** contains the restraint system parameters and thus reflects the environment's status, which is highly influencing the RL agent perception and learning. During the training or evaluation process, a step function chooses actions from the action space. This ensures that the selected action adheres to the defined action space constraints and the state variables are properly updated. The actions which are selected highly depend on the policy of the RL agent.The **reward function** is specified as a conditional statement. Only if the RA falls within a desired interval the weighted sum of the squared reciprocals of the relative loads to the HIC15 and the a3ms legal limit (Gonter et al., 2021) is calculated. If RA is outside the range, a negative penalty discourages actions leading to undesirable states, motivating the agent to avoid them. This incentivizes the agent to minimize HIC15 and a3ms, while trying to keep RA within a reasonable range. The computing time for each step the agent performs either during training or evaluation subsequently only takes a few milliseconds due to the use of the prediction model.

3.2.3. Training process

The RL agent is trained to optimize the loads on the H3 for 300 episodes each containing 500 steps. The PPO algorithm is updated periodically every four episodes to ensure that the agent learns from its

https://doi.org/10.1017/pds.2024.216 Published online by Cambridge University Press

experiences efficiently. This means that after every four episodes of interaction with the environment, the PPO algorithm will perform an update to improve its policy. The mean reward over training episodes is displayed in Figure 3. At the outset, the values indicate approximately 1.7, followed by a sharp ascent to 2.2 within 50 episodes, and convergence becomes discernible after roughly 100 episodes at around 2.3. The training process took 50 minutes on a workstation laptop with 8 CPUs.

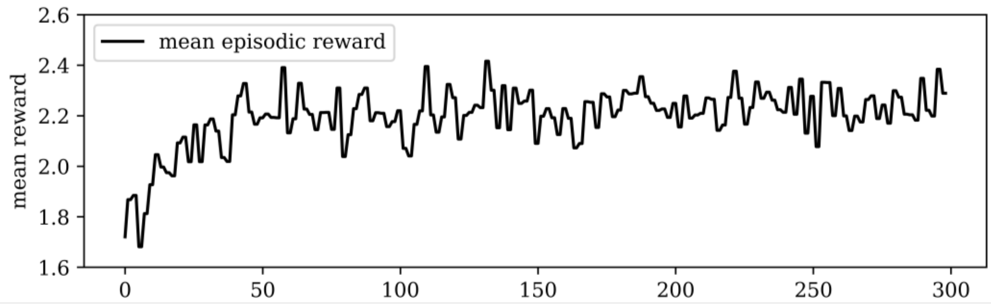

Figure 3. Mean reward per episode during training process of the RL agent

3.2.4. Evaluation and policy-transfer

This section focusses on the evaluation of the best policy found during training on two new initialized environments. Consequently, a newly initialized environment that simulates the H3 is subjected to the learned policy. The evolution of the ATD loads during evaluation is depicted in Figure 4a. Notably, an optimal solution is rapidly reached within only eight steps. While minimal adjustments (9 % less compared to starting value) are needed for the initial chest acceleration value, because the initialized value is already sufficient, the HIC15 is reduced by 42 % so that RA is still over the desired limit marked in Figure 4a. Subsequently, the trained policy can specify a restraint system configuration so that all three values are exploiting the desired ranges as best as possible. Due to the high non-linearities in the system, configuration choices in this context are made using either expert knowledge or conventional optimization techniques. However, local, as well as global optimizers, such as gradient methods or the differential evolution algorithm as used by Horii (2017), are not able to transfer knowledge from one to another problem statement, as it is done in this following step with the RL agent. Here, the best policy trained with the H3 is applied to an entirely unfamiliar environment that emulates the behavior of the HF. Figure 4b displays the ATD loads over the performed steps. The initial values and restraint system parameters themselves are identical for both ATDs due to the same seating position. However, for simplification of this use case it is assumed, that the optimization can be done independently for both ATDs. It is important to note that the response of the HF differs significantly due to variations in seating position within the vehicle, differences in dummy proportions and weight. An optimal solution is still achieved by the agent within ten steps, which is only two more steps than for the H3. Notably, in this scenario, the chest acceleration is reduced by 13 % instead of 9 % compared to the H3. In contrast, HIC15 is reduced by 29 %. RA is also minimized, but such that it remains above the specified safety threshold. To summarize, the RL agent policy learnt on the H3 can be used to either optimize an arbitrary initial state of the H3, but can also be applied to optimize the HF. The slight fluctuations observed in Figure 4 can be attributed to the inherent stochastic nature of the PPO policy and possible non linearities in the response of the tree-based prediction model.

3.2.5. Benchmark with a differential evolution algorithm

For a better contextualization of the results and the performance of the presented RL approach with respect to the reviewed literature in Section 2.2, the conventional global optimization algorithm Differential Evolution (DE), as used by Xue et al. (2014) and Horii (2017), was applied to the identical prediction model for optimizing the ATD loads. The results are depicted in Figure 4c for the H3 scenario. The most important aspect in terms of occupant safety that all values are within desired ranges is fulfilled in both cases. While the chest acceleration for the H3 is identical for both solutions, the DE solution offers a 6 % lower HIC15 with a 21 % higher RA. However, the DE begins the optimization from a better starting point defined by the initial population specified with the Latin Hypercube sampling and

https://doi.org/10.1017/pds.2024.216 Published online by Cambridge University Press

requires significantly more steps to identify the configuration. Convergence becomes apparent at approximately 2000 steps, which is over hundred times later than the trained RL agent. Relative to the starting point the RL agent ensures more load reduction of up to 42 %. Assuming that the global optimizer has found a global optimum, the RL Agent was thus capable of approaching it closely.
Without using a prediction model for this study, each iteration would correspond to a FE simulation, which requires up to 20 hours computing time on a high-performance computing cluster. For such a scenario, the DE algorithm is entirely unsuitable, and smaller models, as described by Xue et al. (2015) and Horii (2017), would need to be employed.

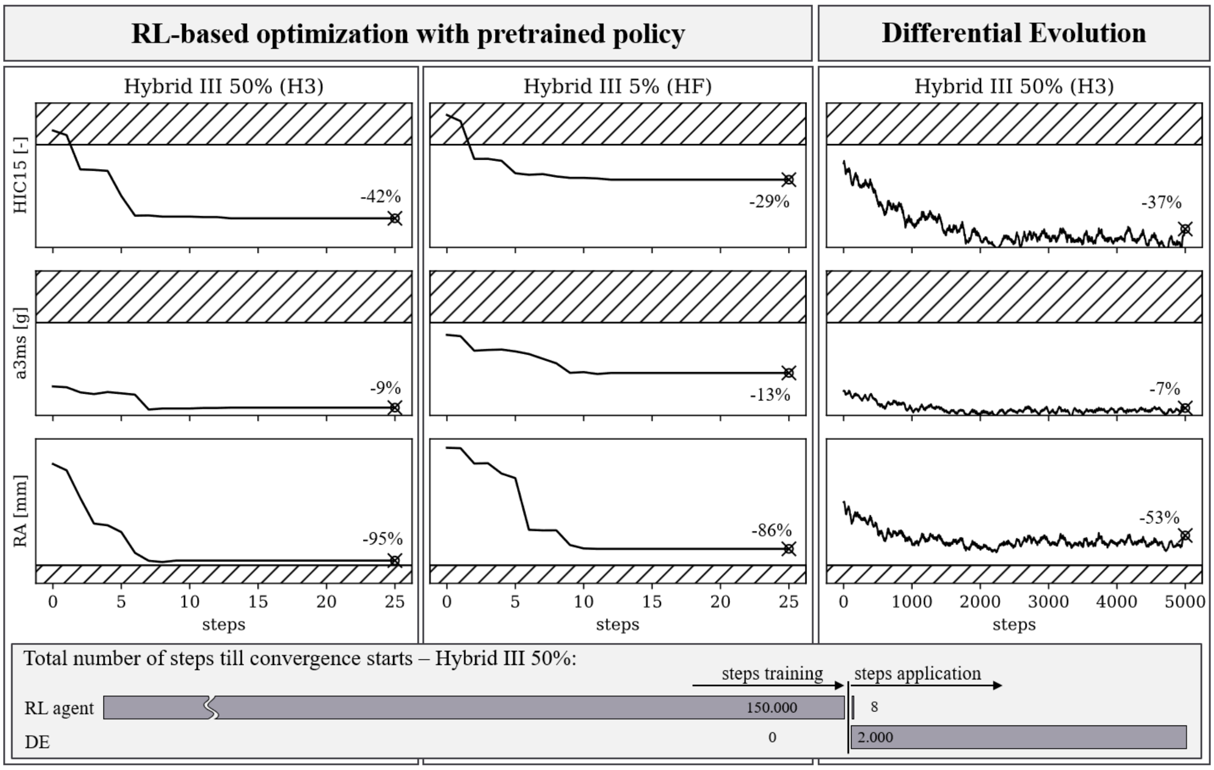

Figure 4. Optimization results of a) H3 and b) HF and c) the reference optimizer

4. Discussion

This chapter discusses the results obtained by the previously presented training and evaluation of the proposed methodology for RL-based restraint system design. Based on the formulated objectives in Section 2.3 the following answers to the research questions can be specified.

1. The RA agent can effectively acquire the capability to minimize ATD loads in a frontal impact crash scenario by making design changes to the restraint system. During the training visualized in Figure 3, the environment is reset for every episode, which means the agent can optimize the system behavior from multiple starting points while the return is maximized.
2. The best policy specified during training can optimize a given start state for both ATDs such that the loads are automatically optimized as depicted in Figure 4. Even though the parameters are identical, which allows for a direct transfer of the policy, the system responses are significantly different due to the weight and the seating position of the HF compared to the H3.
3. The DE approach used as a benchmark since similar ones also have been used in literature (Xue et al., 2014; Horii, 2017), needs significantly longer to converge than the trained RL agent for both problem statements, even though the initial starting point is better. The solution achieved with the DE is slightly better than the one specified with the RL agent. However, both solutions are reasonable from a restraint system development engineer point of view and comply with the desired limits. Regarding computationally expensive FE simulations the required calls of the

https://doi.org/10.1017/pds.2024.216 Published online by Cambridge University Press

prediction model - one in each step - are a suitable measurement, since one call corresponds to one FE simulation. When comparing the steps needed for the optimization if the trained RL agent is used with the steps needed if the DE algorithm is used, the RL agent is over hundred times faster. However, during the training process, the RL agent required 300 episodes each containing 500 steps, which results in 150.000 prediction model calls. Consequently, the total number of steps is larger, even if both problem statements are included when assuming that the optimization of HF would require a similar number of steps like the H3. For practical use cases, the engineers thus have to ensure, that the trained RL agent is used various times. Otherwise, the global effort for the RL agent is significantly larger compared to the conventional optimizer.

Compared to the approaches used in the reviewed literature concerning RL for structural optimization, the agent was directly calling the FE-model (Hayashi and Ohsaki, 2020; Brown et al., 2022; Hayashi and Ohsaki, 2022; Trilling et al., 2022). This works for small academic models, but not for industry standard models as used for occupant safety simulations in modern vehicle development. Subsequently, the training of the RL agent in this scenario would have taken over 200 years. Since the training cannot be reasonable parallelized, a reduction via the use of more computing power is not realizable. However, simulations serving to train a tree-based prediction model can be run in parallel. Furthermore, the model can interpolate between samples with good precision, thus the number of simulations and time can be reduced drastically. Analogous to this work, Trilling et al. (2022) used PPO for the RL agent, which is a policy-based reinforcement learning algorithm. It directly learns the optimal policy, which is a mapping from states to actions. Hayashi and Ohsaki (2020 and 2022) and Brown et al. (2022) have used Q-learning algorithms for their work. These algorithms are value-based and aim to learn the optimal action-value (Q-value) function, which represents the expected cumulative reward of taking a specific action in a given state and following an optimal policy. Beysolow II (2019) considers PPO to be more sample efficient, making it suitable for environments with high-dimensional action spaces or complex dynamics. Hayashi and Ohsaki (2020 and 2022) as well as Brown et al. (2022) also noted that the training of the agent is computationally expensive and the subsequent application of the trained policy to solve similar problem statements is highly efficient.

5. Summary and future work

This study presented a novel approach for the data-driven design of restraint systems in the development of passive safety systems in vehicle development. A significant shortcoming of conventional methods is the limited generalization capabilities from one problem statement to another, which displays significant potential in reoccurring design tasks when developing a variety of vehicle series over multiple generations. To demonstrate the effectiveness of the RL approach in restraint system design in frontal impact crash tests, a tree-based prediction model was trained based on industry standard FE simulations, which captures functional relationships between restraint systems and ATD loads. This reduced the training time for the PPO agent to approximately 50 minutes. To illustrate the agent's generalization capabilities, the policy trained on the H3 was applied to the HF on the same seating position, which has a significantly different structural response. In both cases, rapid convergence was observed in around ten steps. To contextualize these results, a DE algorithm was employed, which required substantially more steps to converge compared to the trained agent. Drawbacks of the RL approach are the considerable investment in the training process as well as the slightly suboptimal solution obtained in this scenario. As also illustrated in the discussed literature, the agent requires numerous training steps, globally even surpassing those of the DE algorithm for the present problem statement which would result in more computationally expensive FE simulations. Therefore, the utilization of the RL agent proves advantageous when the trained policy finds multiple applications in development.

In contemporary research, the application of RL-based restraint system design emerges as a transformative approach. Further research in the field can be the application of this approach to other seating positions, also for further analyzing generalization capabilities. In addition, the focus on structural optimization displays large potentials. Further research for the efficient training process of RL agents within the use of industry-standard large-scale FE simulations is also promising.

https://doi.org/10.1017/pds.2024.216 Published online by Cambridge University Press

References

Beyer, F., Schneider, D. and Schumacher, A. (2021), “Finding three-dimensional layouts for crashworthiness load cases using the graph and heuristic based topology optimization”, *Structural and Multidisciplinary Optimization*, Vol. 63, pp. 1-15. https://doi.org/10.1007/s00158-020-02768-0

Beysolow II, T. (2019), *Applied Reinforcement Learning with Python*, Apress, New York. https://doi.org/10.1007/978-1-4842-5127-0

Brown, N., Garland, A., Fadel, G. and Li, G. (2022), “Deep reinforcement learning for engineering design through topology optimization of elementally discretized design domains”, *Materials & Design*, Vol. 218, pp. 110672. https://doi.org/10.1016/j.matdes.2022.110672

Gonter, M., Knoll, P., Leschke, A., Seiffert U. and Weinert, F. (2021), “Fahrzeugsicherheit”, In: Pischinger, S. and Seiffert, U. (Ed.), *Vieweg Handbuch Kraftfahrzeugtechnik. ATZ/MTZ-Fachbuch*, Springer Vieweg, Wiesbaden, pp. 1073-1160. https://doi.org/10.1007/978-3-658-25557-2_9

Hayashi, K. and Ohsaki, M. (2020), “Reinforcement learning for optimum design of a plane frame under static loads”, *Engineering with Computers*, Vol. 37, pp. 1999–2011. https://doi.org/10.1007/s00366-019-00926-7

Hayashi, K. and Ohsaki, M. (2022), “Graph-based reinforcement learning for discrete cross-section optimization of planar steel frames”, *Advanced Engineering Informatics*, Vol. 51, pp. 101512. https://doi.org/10.1016/j.aei.2021.101512.

Horii, H. (2017), “Multi-objective Optimization of Vehicle Occupant Restraint System by Using Evolutionary Algorithm with Response Surface Model”, *International Journal of Computational Methods and Experimental Measurements*, Vol. 5, No. 2, pp. 163-173. https://doi.org/10.2495/CMEM-V5-N2-163-170

Huang, Y., Zhou, Q., Zhang, X. and Wang, C. (2015), “Optimization study of occupant restraint system concerning variations in occupant size and crash severity in frontal collisions”, *International Journal of Vehicle Safety*, Vol. 8. https://doi.org/10.1504/IJVS.2015.074373

Joodaki, H., Gepner, B. and Kerrigan, J. (2021), “Leveraging machine learning for predicting human body model response in restraint design simulations”, *Computer Methods in Biomechanics and Biomedical Engineering*, Vol. 24, No. 6, pp. 597-611. https://doi.org/10.1080/10255842.2020.1841754

Kracker, D., Dhanasekaran, R. K., Schumacher, A. and Garcke, J. (2023), “Method for automated detection of outliers in crash simulations”, *International Journal of Crashworthiness*, Vol. 28, No. 1, pp. 96–107. https://doi.org/10.1080/13588265.2022.2074634

Langen, T., Falk, K., and Mansouri, M. (2022), “A Systems Thinking Approach to Data-Driven Product Development”, *Proceedings of the Design Society, Vol. 2: DESIGN2022, May* 2022, The Design Society, Glasgow, pp. 1915-1924. https://doi.org/10.1017/pds.2022.194

Muhl, P., Rudolf, S. (2022), “Data-Driven Calibration of Thermal Management Control Systems of an Electric Sports Car”, *31. Aachen Colloquium Sustainable Mobility, Aachen, 2022*.

Rabus, M., Belaid, M. K., Maurer, S., A. and Hiermaier, S. (2022), “Development of a model for the prediction of occupant loads in vehicle crashes: introduction of the Real Occupant Load Criterion for Prediction (ROLCp)”, *Automotive and Engine Technology*, Vol. 7, No. 3–4, pp. 229–244, Dec. 2022. https://doi.org/10.1007/s41104-022-00111-x

Rudolf, T., Schürmann, T., Skull, M., Schwab, S. and Hohmann, S. (2022), “Data-Driven Automotive Development: Federated Reinforcement Learning for Calibration and Control”, *22. Internationales Stuttgarter Symposium*, Springer Fachmedien, Wiesbaden, pp. 369–384. https://doi.org/10.1007/978-3-658-37009-1_26

Schuhmacher, A. (2020), *Optimierung mechanischer Strukturen*, (3rd ed.), Springer Vieweg Berlin Heidelberg, Berlin. https://doi.org/10.1007/978-3-662-60328-4

Sutton, S. and Barto A. (2018), *Reinforcement Learning: An Introduction*, (2nd ed.), A Bradford Book, Cambridge.

Thiele, M., Mullerschön, H., van den Hove, M. and Mlekusch, B. (2006), “Optimization of an Adaptive Restraint System Using LS-OPT and Visual Exploration of the Design Space Using D-SPEX”, *9th International LS-DYNA Users Conference, Detroit*, 2006.

Trilling, J., Schumacher, A. and Zhou, M. (2022), “Einsatz von Reinforcement Learning zur lokalen Versteifung von Extrusionsprofilen in Crashlastfällen”, *Zeitschrift für numerische Simulation und angrenzende Gebiete, NAFEMS-Online-Magazin*, Vol. 62, pp. 59-70, Feb. 2022.

Trilling, J., Schuhmacher, A., Zhou, M. (2022), “Einsatz von Reinforcement Learning zur lokalen Versteigung von Extrusionsprofilen in Crashlaställen”, NAFEMS Magazin, Vol. 61,

Xue, Z., Parashar, S., Li, G. and Fu, Y. (2012), “Optimization Strategies to Explore Multiple Optimal Solutions and Its Application to Restraint System Design”, SAE *International Journal of Passenger Cars - Mechanical Systems*, Vol. 5, No. 1, pp. 540–551. https://doi.org/10.4271/2012-01-0578