Research project for the Research Master Data Science

Project title: 1 Decision support in production and logistics using reinforcement learning

Project overview

Number of students	1
Project Type	Study project
Project Owner	Prof. Dr.-Ing. Christian Schwede
Project Context	Project in collaboration with the supervising professor in preparation for possible consortium research in this context.

Abstract

Simulation models are increasingly being used to plan complex production and logistics systems in order to be able to evaluate the effects of decisions (e.g. resource plans, production programs, bottleneck measures) in detail. In the project described here, reinforcement learning is used to automate the creation of decision scenarios using artificial intelligence. The machine learning process can use the simulation as a training environment.

Short description

Simulation models are increasingly being used to plan complex production and logistics systems in order to evaluate the effects of decisions (e.g. resource plans, production programs, bottleneck measures) in detail. These models have recently been linked to digital real-time images of the production and logistics systems, so-called "digital twins". This means that the models can also be used for short-term planning and real-time control. If, for example, an important delivery arrives late at the factory, the effects can be evaluated ad-hoc using the simulation based on the real-time data of the digital twin and countermeasures can be checked.

In addition to the time-consuming creation of simulation models, the decision-making process is still largely manual. Decision scenarios are created manually in the decision support system, evaluated by the simulation and then manually adjusted again.

In the project described here, reinforcement learning is used to automate the creation of decision scenarios using artificial intelligence. The machine learning process can use the simulation as a training environment. It generates input and receives feedback on its quality through the simulation. In this way, decision-makers can be supported by intelligent suggestions and less serious decisions can be made completely autonomously.

Task definition

The student will create a simulation model for a given production/logistics system using Siemens Plant Simulation software. This model will serve as the basis for the development of a reinforcement learning process that is designed to generate optimal decision scenarios for specific framework conditions.

Reference to the topic of data science

Reinforcement learning methods are a core topic of data science and are also covered in the courses of the research master's program.

Available resources

- Information required to create the simulation model (system description, throughput times, capacities, etc.) is provided by the supervising professor
- Siemens Plant Simulation is provided in the context of the AI Lab at Bielefeld University of Applied Sciences
- Hardware for reinforcement learning is available from the Data Science Lab or the CfADS at Bielefeld UAS

Project plan

First semester: formulation of the research exposé, familiarization with Plant Simulation, collection of the information required to create the simulation model

Second semester: Creation of a simple simulation model, research on relevant work in the field of coupling simulation and machine learning

Third semester: Development of initial machine learning methods, development of an interface to Plant Simulation, evaluation of initial results

Fourth semester: Creation of a complex simulation model, implementation and comparison of additional methods, fine-tuning and optimization of the methods, improvement of the results, final evaluation

Necessary competencies

Mandatory:

- Programming skills

Optional:

- Simulation experience
- Experience with production and logistics systems

Acquirable competencies

- Use of reinforcement learning methods
- Discrete-event material flow simulation with Plant Simulation
- Planning of production and logistics systems

Research project for the Research Master Data Science

Project title: 2 Learning digital twins

Project overview

Number of students	1-3
Project Type	Project with external partners
Project Owner	Prof. Dr.-Ing. Christian Schwede
Project Context	Project in cooperation with the **Fraunhofer Institute for Software and Systems Engineering** (ISST) in Dortmund and in the context of the **Centre for Applied Data Science** (CfADS) and **the Institute for Data Science Solutions** (IDaS); employment as a student assistant at Fraunhofer is possible

Abstract

The aim of the project is to learn simulation-based digital twins for planning and controlling production and logistics environments from IoT data. It builds on and extends the open source framework OFacT developed by HSBI and Fraunhofer ISST. The data comes from Fraunhofer customer projects and the CfADS IoT Factory. In addition to process mining methods, generative models such as large language models (LLMs) are also used.

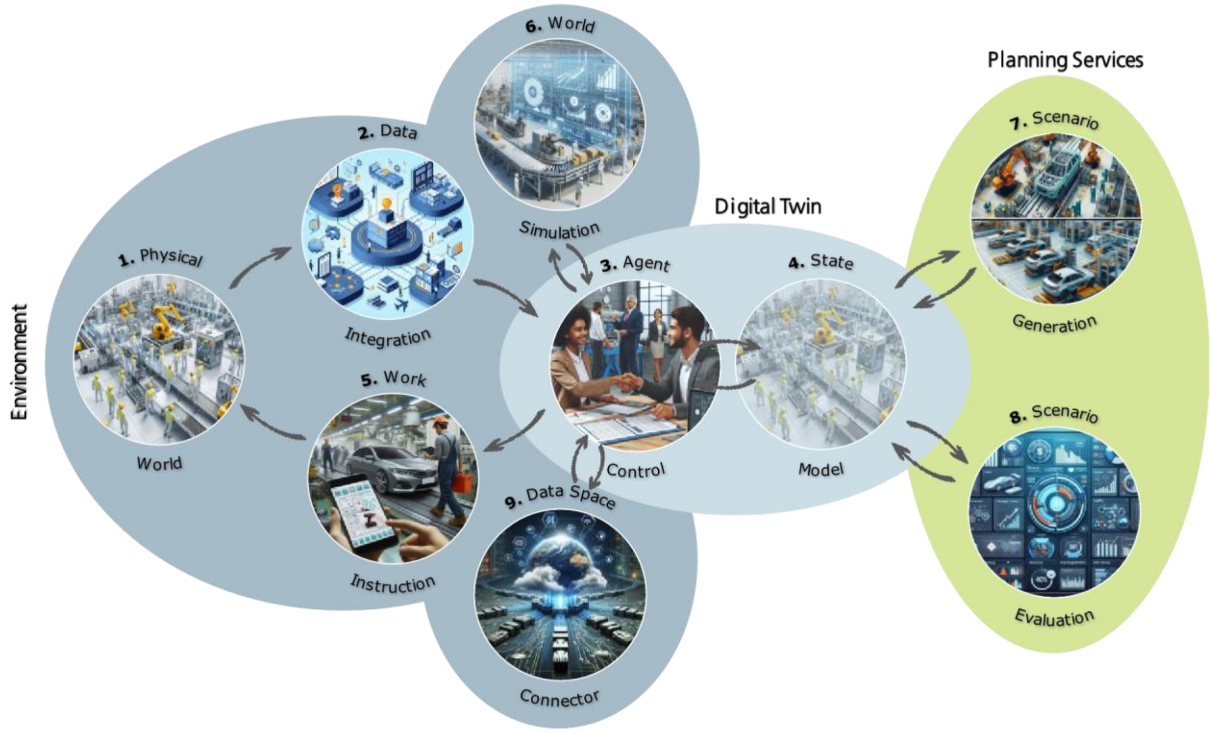
Short description

Digital twins are the key to the optimal design, planning and control of production and logistics environments. These software frameworks allow access to a constantly updated digital image of the system, enable the optimization and simulation of scenarios in this virtual environment and communicate changes directly back to the store floor by means of work instructions. In this way, decisions can be evaluated and optimized in all phases of the system's life cycle before they are made.

Both long-term changes to the system and reactions to short-term disruptions can be optimally implemented. The potential for companies is estimated to be very high (forecast market volume of USD 10 billion in 2028[1]). Fraunhofer ISST and HSBI have created the open source framework OFact (Open Factory Twin), which allows small and medium-sized companies in particular to digitize their production systems.

OFacT
Open Factory Twin

OFacT is based on a standardized state model that can be used to map any discrete material flow system. Any data interfaces can be addressed via the data integration module. The system is controlled by a multi-agent system so that control rules can be mapped and changed in a highly flexible manner and separately from the state model. The agents also control the material flow system by means of work instructions.



[1] Bain & Company: Global Machinery & Equipment Report 2024

As an environment, the framework can draw on a simulated reality to evaluate scenarios, for example. Planning services allow the creation and optimization of alternative courses of action and a dashboard allows evaluation based on standard key figures.

An important hurdle in the implementation of digital twins in small and medium-sized companies is the cost of manual modelling by experts. These costs are incurred not only during the creation, but also during regular updates. The aim of this project is therefore to learn digital twins from company data.

Task definition

In the project, the student should use methods from the field of machine learning to learn OFacT models based on data from production environments. The IoT data or event logs from the store floor of companies often already contain a lot of the graded information. Data from existing industrial projects from Fraunhofer ISST or the IoT Factory on the Gütersloh campus can be used. In addition to process mining methods, regression and classification methods, e.g. based on deep neural networks, will be used. It will also be investigated to what extent parts of the models can also be created with current Large Language Models (LLM). For this purpose, it may be useful to use Retrieval-Augmented Generation (RAG) to provide the LLM with domain-specific information.

Reference to the topic of data science

The regression and classification methods used, as well as LLM and RAG, are part of the Research Master's course and core areas of Data Science.

Available resources

- Data is provided via Fraunhofer ISST and the IoT Factory
- The open source framework OFacT is being developed at HSBI and is available
- Hardware for machine learning is available via the Data Science Lab, CfADS and the AI computing cluster yourAI at HSBI

Project plan

First semester: Preparation of a research exposé as an examination. Familiarization with the task and the OFacT framework.

Second semester: Literature research on the generation of models using machine learning methods, in particular LLMs and RAG. Review of data sets and application of first simple methods of process mining and existing language models. Preparation of a paper that gives an overview of the research area is an examination achievement.

Third semester: Practical implementation of your own solution and generation of an OFacT model for a specific application. Publication of a paper with the first results is an examination achievement.

Fourth semester: Master's thesis and colloquium. Final evaluation by comparing different methods.

Necessary competencies

Mandatory:

- Programming skills (e.g. Python) Optional:

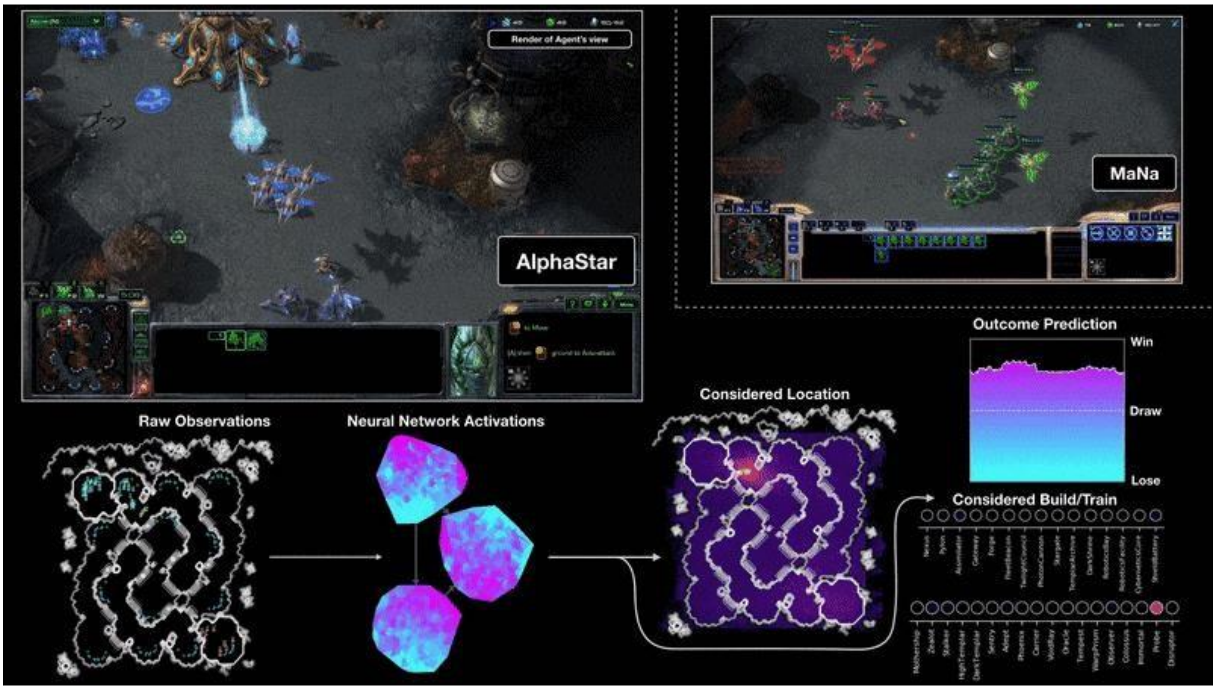
- Experience with machine learning
- Experience with LLMs

Acquirable competencies

- Development and use of digital twins
- Generation of models using LLMs and RAGs
- Publication of scientific papers at international conferences
- Competencies in working in a team - Open source development

Research project for the Research Master Data Science

Project title: 3 Interpretable and transferable AI for StarCraft2



Project overview

Number of students	1-3
Project Type	Study project
Project Owner	Prof. Dr. Ing. Christian Schwede
Project Context	The study project is based at the Digital Lab on the Gütersloh campus. The team consists of a research assistant and two other students.

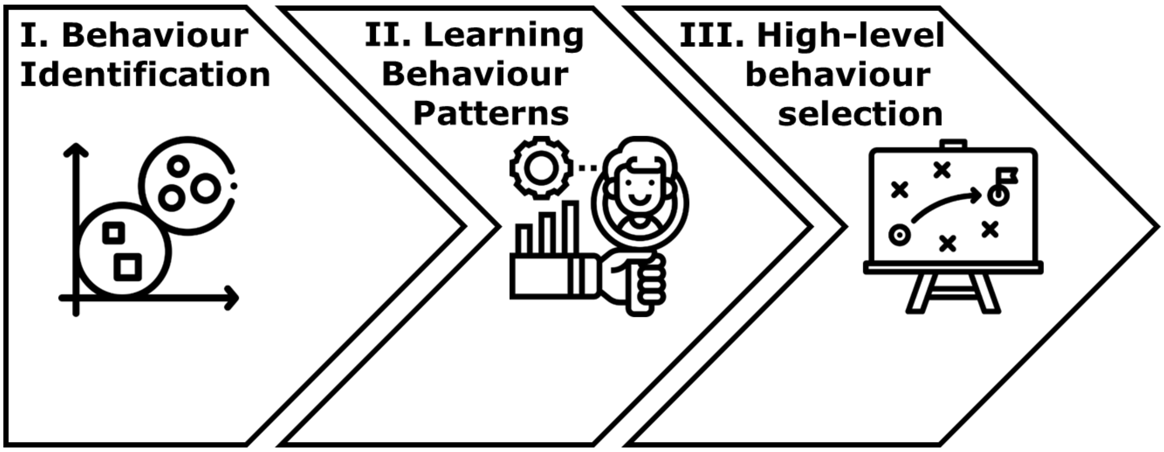
Abstract

In this project, cutting-edge deep reinforcement learning methods are being used to work on an artificial intelligence for StarCraft2 that makes knowledge preservable and transferable and is also comprehensible for humans (explainability).

Short description

Computer games have been an exciting challenge for the development of intelligent agents since the early days of artificial intelligence. Major breakthroughs in AI research were often achieved through the development of programs that were able to defeat human opponents. Deep Blue beat the reigning world champion Kasparov in chess in 1996, IBM's Watson won "Jeopardy!" against two against two professionals in 2011 and Google's AlphaGo beat Lee Sedol, one of the world's best Go players, in 2016. Deepmind's AlphaStar has been playing in the various leagues of Starcraft2 since 2019 and now beats 99.8% of players. Even if this AI is unsurpassed in terms of performance, it hardly fulfils two other important requirements for AI: transferability/learning efficiency and explainability. On the way to strong AI, learning efficiency, transferability and not forgetting knowledge is crucial, while for us humans, insight into the what and why of AI's actions is of crucial importance.

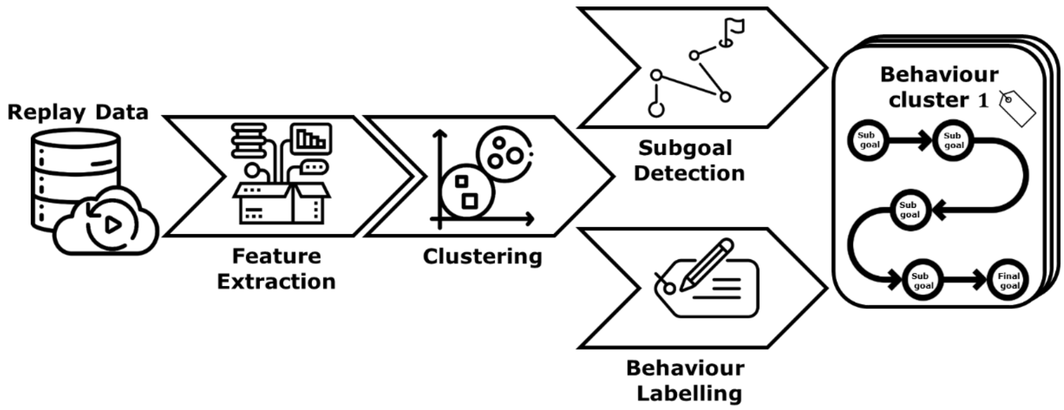
This project is therefore working on an artificial intelligence for StarCraft2 that fulfils these two requirements. The basic idea is to develop an AI that learns to select a suitable action module (strategy) depending on the situation and not just a single action. In this way, knowledge can be retained in the building blocks and complex actions can be made more explainable by encapsulating them in strategies. The individual strategy modules should first be recognized and then learned on the basis of example games. The additional construction of a subgoal graph, which puts the intermediate goals of individual strategy modules in order, makes them even more explainable for humans. The entire approach is divided into three sub-steps and shown in the following diagram. First, strategy clusters are learned from example data using time series clustering, then strategy modules are learned on the basis of these clusters using supervised or reinforcement learning and finally the high-level strategy is learned using hierarchical reinforcement learning.



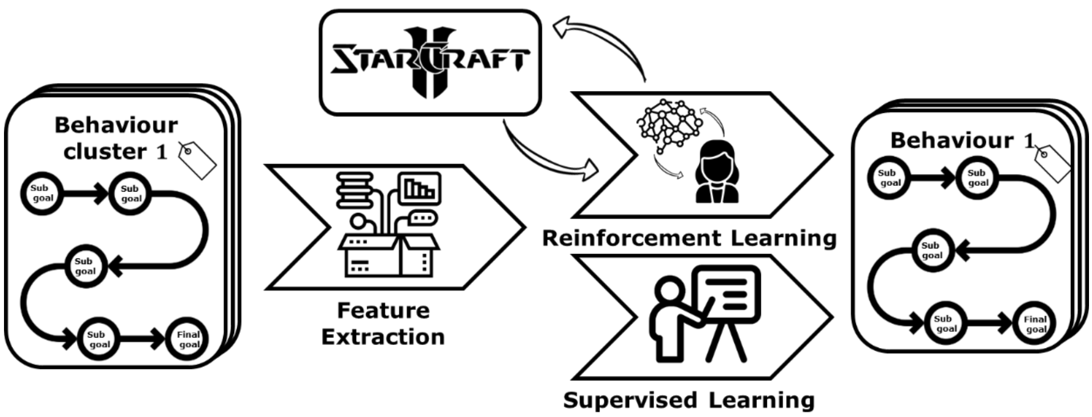
Task definition

Depending on their interests, students can work in one of the sub-areas described above. The focus of the work can relate to one of the three process groups mentioned.

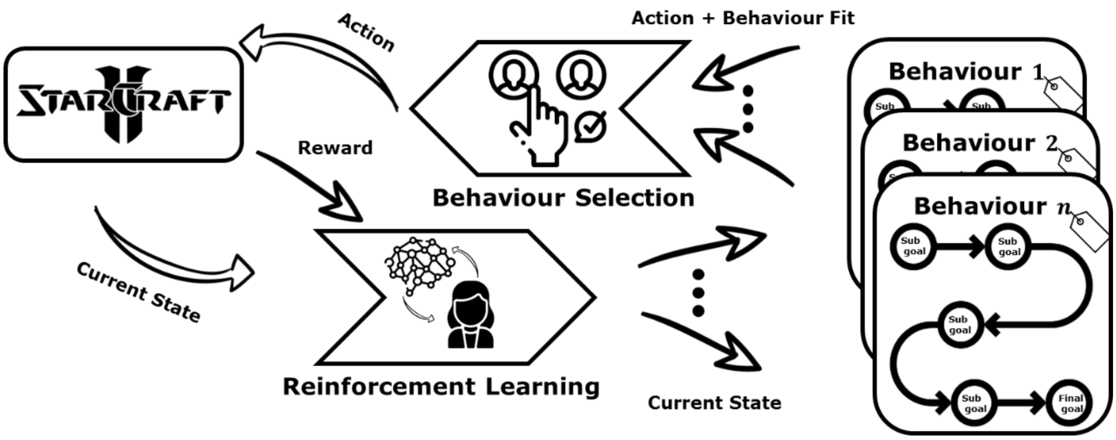
The first area of strategy recognition will mainly work with clustering methods. Tasks of feature extraction in the variable and high-dimensional input space of StarCraft2, the selection of suitable distance metrics, the selection and implementation of clustering methods for variable time series, the development of methods for the recognition of sub-targets and automatic labelling can be worked on.



The second area can include either supervised or deep reinforcement learning methods. Supervised methods are used to imitate the behaviour of the clustered strategies, while reinforcement learning is intended to learn its own behaviour based on the subgoals.



The third area comprises hierarchical reinforcement learning methods. This involves learning to select the appropriate strategy module for the respective situation. In addition, the selection behaviour must be influenced in such a way that the strategy is not selected anew after each action.



Reference to the topic of data science

Clustering, supervised and reinforcement learning methods are core topics in data science and are also covered in the research master's courses.

Available resources

- Millions of replay data and a Python API for StarCraft2 are available
- Hardware for machine learning is available from the Data Science Lab or the CfADS at Bielefeld UAS

Project plan

First semester: Formulation of the research exposé, familiarization with the game, the source code and the API.

Second semester: Development of a simple algorithm in the respective field for a simple game scenario, research on relevant works in the subject area

Third semester: Development of first more complex procedures for simpler and more complex game scenarios, evaluation of first results

Fourth semester: Creation of a complex scenario, implementation and comparison of further procedures, fine-tuning and optimization of the procedures, improvement of the results, final evaluation

Necessary competencies

Mandatory:

- Programming skills
- Enthusiasm for games/ StarCraft2 and AI

Optional:

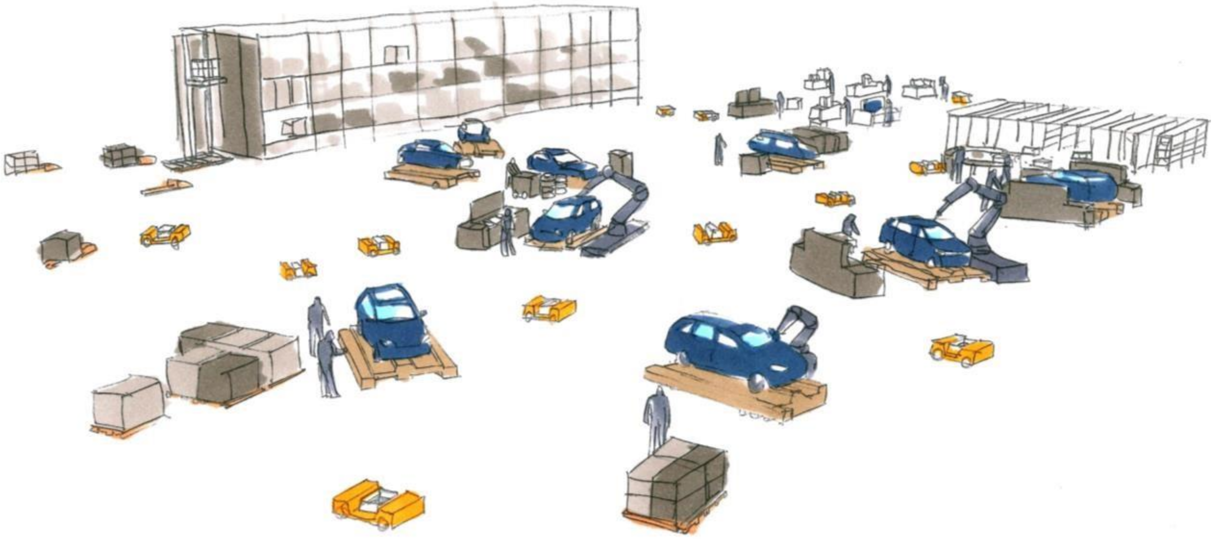
- Experience in ML

Acquirable competencies

- Use and (further) development of state-of-the-art machine learning methods in highly complex environments

Research project for the Research Master Data Science

Project title: 4 AI for highly flexible, multi-variant mass production



Project overview

Number of students	1
Project Type	Project with external partners
Project Owner	Prof. Dr. Ing. Christian Schwede
Project Context	Project in cooperation with the **Fraunhofer Institute for Software and Systems Engineering (ISST)** in Dortmund and in the context of the **Center for Applied Data Science (CfADS)** and the **Institute for Data Science Solutions (IDaS)**; employment as a student assistant at Fraunhofer is possible;

Abstract

The constant individualization of products poses challenges not only for German automotive production, forcing a complete rethink of production organization. Away from clocked flow production towards decentralized matrix production systems. Here, swarms of autonomous robots coordinate the value creation processes and adapt flexibly to the current situation. While the basic idea of the concept is clearly outlined, there are still many open research questions in detail, particularly concerning the behaviour of the individual robots. This is where the project comes in and explores different AI strategies in a virtual testbed.

Short description

In matrix production, production is coordinated decentral by software agents in a multi-agent system (MAS). The resource agents negotiate with the order agent, for example, about the provision of value added processes that are necessary for the realization of the order. The freedom of the sequence of production steps is only limited by the technical properties of the product. In this way, the entire system can adapt flexibly to unplanned incidents, such as a delivery bottleneck or a machine breakdown. While the concept was initially met with rejection and scepticism in the automotive industry, intensive efforts have been made in recent years to research and promote its operationalization. In principle, there are still many unanswered questions regarding its efficient use that need to be researched.

The open source OFacT (Open Factory Twin) software framework has been developed by the HSBI research group in cooperation with the Fraunhofer ISST in Dortmund for several years. Among other things, this tool allows highly flexible matrix production to be modelled using agent-based simulation and various strategies and design variants to be dynamically evaluated. The framework can also be used as a digital twin and can be connected directly to production via sensor data streams and used to control operational processes.

Task definition

The student has the opportunity to work on one of the following topics in order to explore the possibilities and limitations of matrix production and to contribute new solution concepts to the realization of the state of the art.

Agent control for decentralized production

One of the most crucial open questions is the optimal control of the agents. In particular, the synchronization of material flows with the main product is a decisive factor in preventing waiting times and maintaining flexibility at the same time. The development and implementation of efficient control strategies for the MAS is the core task here, which includes the question of whether and to what extent planning and reservations should be made in the future. Methods of optimization, search and planning algorithms from AI or reinforcement learning can be used here.

Layout design and optimization

Even if the term matrix production implies a grid-like arrangement of the production stations, the question of the optimal arrangement of the various elements of matrix production (stations, storage locations for parts, loading stations for transport vehicles, parking spaces for intermediate buffering of car bodies, etc.) has not yet been finally decided. Optimization algorithms, which are used in combination with simulation as an evaluation method, must be developed here in order to determine the optimal layout for a specific production setup. The question of how strongly the optimum layout depends on the specific order mix is relevant and could imply the need for regular re-planning of the layout.

Routing and congestion analysis

The formation of traffic jams in front of the production stations or on the various routes is a central aspect of the efficiency of the overall system. For this purpose, the simulation environment must be expanded to include a route planning element that also recognizes temporary route conflicts. AI search algorithms from the area of route finding must be implemented and various strategies tested. Finally, existing control strategies and layouts must be examined and evaluated against the background of congestion in the system.

Acceleration of negotiations by means of reinforcement learning

A weak point of multi-agent systems is the amount of communication that arises from the decentralized negotiations between the agents. Negotiation strategies need to be examined, implemented and compared and, in particular, unnecessary communication needs to be avoided.

Reinforcement learning can be used to recognize when requests from agents are not expedient and can therefore be avoided in advance.

Agent-based value creation networks

The integration of logistical processes from the value creation network around production is another open topic. In addition to the development of suitable supply and disposal strategies, which must meet the needs of flexible matrix production, the visionary question of whether decentralized, autonomous value creation can also be extended beyond the factory boundaries should be answered. To this end, the agents must negotiate contracts with suppliers and service providers on the basis of money as a means of payment. The corresponding negotiation algorithms and scenarios must be implemented and evaluated.

Reference to the topic of data science

Software agents and multi-agent systems, search algorithms and reinforcement learning are core areas of AI and are addressed in various courses of the research master's program.

Available resources

- The modelling framework OFacT is provided by HSBI and Fraunhofer ISST
- Industry-related scenarios of individual bicycle production exist to validate the developments
- Scenarios from industrial customers of ISST and the IoT Factory in Gütersloh also exist for evaluating individual processes
- Hardware for more complex machine learning is available via the Data Science Lab or the CfADS of the

Project plan

First semester: Formulation of the research proposal, familiarization with the OFacT environment, selection of the topic.

Second semester: Modelling of an initial scenario, research on relevant work in the subject area and preparation of a report on the state of the art.

Third semester: Implementation and evaluation of initial procedures, evaluation of initial results based on the scenario, writing a paper for an (inter)national conference.

Fourth semester: Development of further procedures and optimization of the overall approach, evaluation of the quality and evaluation of the results in the Master's thesis

Necessary competencies

Mandatory:

- Programming skills (ideally in Python)

Optional:

- Experience with agent systems
- Experience with simulation and optimization
- Background knowledge of production logistics and production systems
- Experience with the implementation of reinforcement learning / ML methods

Acquirable competencies

- Use of reinforcement learning methods
- Development and evaluation of MAS
- Development of intelligent agents
- Working with agent simulations
- Planning of production and logistics systems, in particular market production systems and Industry 4.0

Research Project for the Research Master Data Science

Project title: 5 NLP-Driven Configuration for Mass Customization (NLP-KiMaC)

Project overview

Number of Students	1
Project Type	Funded project with external partners
Project Owner	Prof. Dr. Stefan Berlik / Dr. Mohammad Seidpisheh
Project Context	An interdisciplinary research initiative in collaboration with industry leaders, focusing on the implementation of Mass Customization using Natural Language Processing (NLP) techniques. Partners include leading manufacturing companies. Opportunities for further employment as a research assistant may be available.

Abstract

This project aims to redefine production processes through NLP-enhanced Mass Customization, combining the personalization of custom-made products with the efficiency of mass production. By leveraging NLP technologies, the project focuses on enabling intuitive and automated configuration systems that enhance user interaction and streamline production workflows. Key innovations include the development of intelligent assistants for product configuration and manufacturing order

optimization, transforming static and manual configuration workflows into automated, scalable, and user-friendly processes.

Short description

The KiMaC project represents a pioneering initiative in the field of mass customization, combining the expertise of GRAUTHOFF Türengruppe GmbH and Solarlux GmbH with advanced AI technologies to revolutionize the production of doors and glass folding walls. This initiative not only emphasizes technological innovation but also the sustainability and customization capabilities that AI brings to the manufacturing processes. The proposed follow-up project will focus on the integration of NLP techniques into the KiMaC project.

GRAUTHOFF Türengruppe GmbH is a renowned player in the building materials industry, specializing in the manufacturing of doors. With over 60 years of history that began in a small workshop, GRAUTHOFF has grown into one of Germany's most significant and advanced door manufacturers.

Solarlux GmbH has established itself as a leader in the production of glass folding walls, conservatories, and facade solutions for over 35 years. Solarlux is celebrated for its exceptional quality, sustainable practices, and innovative designs.

GRAUTHOFF and Solarlux will contribute real-world production scenarios and provide operational data to support the development and rigorous testing of the NLP-enhanced configuration systems. Their involvement ensures the solutions are practical, effective, and aligned with industry needs. The NLPKiMaC project aims to leverage NLP technologies to streamline the customization process, enabling user-friendly and efficient production of highly personalized products. These intelligent systems facilitate improved communication between stakeholders, enhanced precision in design, and faster production cycles without compromising on quality.

The synergy between GRAUTHOFF's expertise in door manufacturing and Solarlux's innovation in glass solutions, coupled with the transformative potential of NLP, positions NLP-KiMaC to redefine interaction, efficiency, and personalization in manufacturing.

Task definition

Students will play a key role in integrating NLP into configuration systems, creating solutions that transform customization into a seamless and efficient experience. Key tasks include:

1. **Developing NLP-Driven Configuration Systems:**
   Design intuitive systems enabling users to interact naturally with customization platforms through conversational interfaces, simplifying complex workflows.
2. **Building Intelligent Assistants for Configuration:**
   Create NLP-powered assistants to provide real-time suggestions, streamline decision-making, and reduce configuration errors.
3. **Evaluating Integrated Systems:**
   Assess system usability, operational speed, and scalability in real-world production settings to validate their efficiency and impact on customization processes

Reference to the topic of data science

The project integrates core Data Science methodologies, including natural language understanding, machine learning, and predictive modelling. These technologies tackle real-world challenges, including simplifying configuration processes, capturing customer requirements accurately, and enabling efficient user interactions.

Available Resources

Collaboration with industry partners provides access to real production environments, operational data, and IT infrastructures for the development and implementation of AI models and systems. A modern cloud environment will be utilized for software deployment, supported by extensive computational resources for ML training and optimization tasks provided by the academic institution's Data Analytics Cluster.

Project plan

First Semester: Research exposé detailing the design of NLP-driven configuration systems and the initial integration of AI models.

Second Semester: Systematic literature review on NLP applications in Mass Customization and user interaction design.

Third Semester: Development and testing of NLP-enabled configuration systems in simulated environments.

Fourth Semester: Master thesis focusing on the integration, evaluation, and optimization of these systems in real-world production settings, followed by a colloquium.

Necessary Competencies

Mandatory:

- Bachelor's degree with a focus on computer science, data science, or a related field.
- Advanced skills in programming languages such as Python.
- Basic understanding of machine learning, statistics, and optimization.
- Interest in applying NLP and AI to industrial challenges.

Optional:

- Experience with advanced NLP techniques.

Acquirable Competencies

- Expertise in applying NLP technologies to complex, real-world challenges in customization and manufacturing.
- In-depth understanding of Mass Customization workflows using NLP.
- Advanced research skills through the design, development, and evaluation of cutting-edge NLP systems in an interdisciplinary context.
- Enhanced research skills through the development, implementation, and evaluation of cutting-edge AI systems in an interdisciplinary setting.

Research Project for the Research Master Data Science

Project title: 6 Machine learning for detecting anomalies and prediction of interactions based on movement information in a smart home environment

Project overview

Number of Students	1
Project Type	Project with external partners
Project Owner	Prof. Dr.-Ing. Thorsten Jungeblut
Project Context	Project in cooperation with **KogniHome - Technikunterstützes Wohnen für Menschen e.V.** in Bielefeld; Data sets and hardware resources are provided; Employment as a research assistant is possible; There is the possibility of close cooperation with members of KogniHome e.V., including Bethel, Steinel, Hettich, HUM Systems, C&S GmbH and many more.

Abstract

The aim of the project is the co-development of a system that uses AI methods to learn the usage behaviour of the residents of a smart home in order to subsequently provide suitable recommendations for action or, if anomalies are detected, to initiate appropriate measures. The scientific challenge of the project is the application and evaluation of machine learning methods for learning interaction patterns based on the data provided by minimal-sensor technology, as well as the prediction of future interaction and anomaly detection.

Short description

Digitalization is penetrating more and more areas of life, and the associated networking of diverse components in our daily environment is leading to a situation in which constantly higher demands are being placed on everyday technology. Systems need to become smarter, automated and autonomous. Ideally, technology will anticipate the needs of people, and they will no longer need to make corrections. Needs-based human-technology interaction requires the intelligent technical system (ITS) to adapt to the context of use and not the other way round. A requirement for intuitive interaction is therefore first and foremost the reliable recognition of the context of use, i.e. where the user is located and what action they are currently performing. Knowledge of regularly occurring interaction patterns enables the ITS to predict future interactions and predictively control assistance functions. Imaging sensor technology (e.g. cameras, high-resolution time-of-flight sensors) enables powerful recognition of the context of the action through person, object, gesture or even facial recognition, but brings with it the problem of collecting personal data. This may be undesirable for data protection reasons, particularly in private domestic environments, but also in office buildings or production facilities. Minimal sensors, such as motion or presence detectors, do not directly generate personal data, but only record selective information about the presence of people or objects in a spatially restricted area. For comprehensive detection of the context of use, the combination of many simple sensors promises sufficient detection of the context of use while at the same time preserving privacy. If more complex sensor technology is required for an assistance function (e.g. for voice or gesture recognition), it only needs to be activated (and only then) if the use of the assistance system is at least foreseeable. However, the continuous recording of the context of use also makes it possible to learn regular patterns of behaviour. The ITS can derive expected future interactions from these learnt action patterns and make recommendations for the activation of assistance functions or prepare them (predictive control). In addition, deviations from the potential control state (anomalies) can also be recognized and responded to appropriately; the derivation of recommendations for action by the ITS can be rule based in the simplest implementation (e.g. "If the motion detector in the hallway is activated, then the light should be switched on"). In systems with many sensors, however, this is very complex, cannot be flexibly customized and is difficult to scale to growing environments. In addition, different sensors provide different abstract information (e.g. local movement (PIR, ultrasound) or movement across several rooms (HF/microwaves), binary "on/off" or distances) or even more complex information from camera-based systems (e.g. number of recognized people/pets/objects (robot vacuum cleaner/transport platform)). It is therefore necessary to abstract the various pieces of information, anonymize them if necessary and use them in a holistic approach to learn the context of use, estimate future interaction and detect anomalies. The scientific challenge of the project is therefore the application and evaluation of machine learning methods for learning interaction patterns based on the data provided by the minimal-sensor technology, as well as the prediction of future interaction and anomaly detection.

Task definition

In this project, the student will co-develop an ITS that uses machine learning methods to learn the usage behaviour of the residents of a smart home in order to subsequently provide suitable recommendations for action based on the real-time data or, if anomalies are detected, to initiate suitable measures. The development of the concrete application scenario, together with the KogniHome e.V. team, is part of the project. To illustrate this, the following scenario is described: "Sabine (74) gets up every morning between 7 and 8 a.m., goes to the toilet and then makes herself a coffee". Three activities can be recognized in this action sequence: Getting up, going to the toilet, making coffee. Each of these activities can now lead to recommended actions or anomalies. For example, the ITS could prepare for going to the toilet with suitable lighting or switch on the coffee machine. At the same time, the ITS monitors the activities and recognizes deviations from the usual behaviour. For example, if Sabine falls on the way to the toilet and does not get up again on her own, the ITS should recognize this and, for example, call the emergency services or the nursing home.

Reference to the topic of data science

The evaluation and application of AI/ML methods for condition monitoring and prediction are a core topic of data science and are covered, for example, in the modules "Data Mining & Machine Learning" and "Artificial Intelligence". The recording of high-resolution measurement data (e.g. consumption data) from many sensors in complex living environments highly demands on the organization and processing of the data. This is the core of the "Big Data Architectures" module.

Available Resources

- Information required to create the scenario (system description, logistical processes, relevant key figures) is provided by KogniHome e.V.
- Access to the research apartment of KogniHome e.V.
- Extensive test data sets are available via the KogniHome e.V. research apartment - The contact person at KogniHome e.V. will be available for the duration of the project
- Required materials will be provided by KogniHome e.V.
- Hardware for the more complex machine learning is available via the Data Science Lab, CfADS and the AI computing cluster yourAI at HSBI.

Project plan

First semester: Preparation of a research exposé as examination performance. Familiarization with the concepts and structures (IoT software, smart home protocols, interfaces, etc.) of the KogniHome e.V. research apartment.

Second semester: Creation of the system concept for anomaly detection for predictive control of assistance systems. Research on relevant work in the field of the use of AI methods for learning event sequences and for anomaly detection. Preparation of a paper that provides an overview of the respective research area as an examination performance.

Third semester: Practical implementation of various machine learning methods in the field of unsupervised and (semi-)supervised learning to optimize automated interaction between humans and machines, as well as their evaluation.

Fourth semester: Master's thesis and colloquium. Final evaluation by comparing the implemented strategies. Preparation of a paper with initial quantitative results as an examination.

Necessary Competencies

Mandatory:

- Programming skills (especially Python)
- Experience with the version control system "git"

Optional:

- Experience in electronics development

- Experience in the field of smart home technologies/IoT devices

- Programming of microcontrollers

Acquirable Competencies

- Artificial intelligence methods for problem solving

- Sensor-based information processing

- Predictive assistance and anomaly detection (transferability to industrial processes)

- Expertise in working in a team

Research Project for the Research Master Data Science

Project title: 7 Efficient Execution of Object Detection Algorithm on Edge Devices

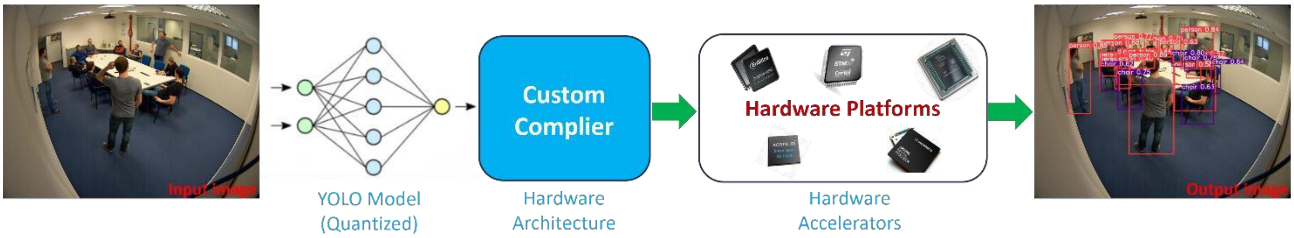


Figure 1: Efficient mapping of YOLO-Vx model on edge devices

Project Overview

Number of Students	1
Project Type	Study project in collaboration with an external partner
Project Owner	Prof. Dr.-Ing. habil. Thorsten Jungeblut Dr.-Ing. Qazi Arbab Ahmed
Project Context	**CareTech OWL** transfer project in collaboration with an external partner, **Steinel GmbH**, as part of the CareTech OWL research project. Parallel employment as a student research assistant (WHK) is possible.

Abstract

The artificial intelligence (AI) paradigm has replaced a handful of traditional computer vision techniques to intelligently predict the correct output of a complex system under tight constraints and varying conditions. However, many AI algorithms require huge amounts of hardware resources due to their massive computational requirements. The main task of this project is the development of a framework for the efficient implementation of object detection algorithms on highly resource constrained hardware architectures (low-end FPGAs, embedded microcontrollers). Based on the framework, automated methods for exploring the design space of suitable combinations of object detection algorithms and hardware in terms of HW/AI-co-design will be explored. In particular, industrial applications with high latency requirements are targeted, such as real-time human presence detection.

Short description

For the efficient execution of AI algorithms, techniques such as federated learning (FL) and cognitive edge computing (CEC) have already shifted the load of training and inference of neural networks from the cloud to the edge. The partitioning of the execution of the application has a significant impact on the performance and resource efficiency of the overall system. For example, a distinction can be made here between fundamentally different approaches of decentralized feature extraction as close as possible to the sensor with subsequent fusion versus central processing in the cloud. The first approach requires powerful edge hardware for preprocessing and potentially offers advantages for high real-time requirements. The second approach places higher demands on the communication infrastructure but potentially enables the execution of more complex networks. The main goal of this project is therefore to explore automated methods for exploring the design space of suitable combinations of AI-based objection detection methods and resource-constrained hardware in terms of HW/AI codesign.

In the area of efficient execution of AI procedures on embedded systems (cognitive edge computing), great progress has been made in the past. At all levels of the different processing concepts (cloud, fog, edge, very edge) in the production chain, a variety of potential hardware architectures and AI accelerators can be found, which differ in the available system resources (e.g., performance or power consumption). Examples of relevant hardware architectures are embedded microcontrollers with integrated AI acceleration, embedded GPUs/FPGAs, dedicated AI hardware accelerators, or high-end GPUs/FPGAs from the HPC domain.

We consider the complete processing chain, starting with the pre-processing close to the sensor (very edge), by connecting the groups of sensors and their information in the edge, via local decentralized cloud instances (fog), to the central station of all relevant information (cloud). At each level in the processing chain, individual approaches exist to locally optimize the resource efficiency of AI methods, for example, by reducing numerical precision (e.g., from 32 to 16, or 8 bits) to enable more efficient execution on specialized hardware or to minimize local storage space. The goal of this project is to determine an optimal combination of hardware and object detection algorithms (e.g., YOLO) as part of a holistic design space exploration close to the sensor.

Task definition

In this project, students will develop a framework for the efficient execution of AI-based object detection algorithms on resource-constrained edge devices. In order to support the sustainable use of the developed methods and frameworks by the industrial partner, universally applicable models and automated design tools are to be made available in this project in the form of a standard development toolkit. The first main challenge is the compression of the latest object detection algorithm (e.g., YOLO) to reduce the model size, yet with acceptable accuracy, using AI approximation techniques. The next challenge is to develop resource-efficient techniques (inference compiler) to map the compressed model to a suitable hardware platform e.g., low-end FPGAs or embedded (AI-) microcontrollers, considering the resource utilization in terms of area, power/energy, and latency/throughput, as demonstrated in Figure 1.

Reference to the topic of data science

The evaluation and application of AI/ML methods in the field of "machine vision", e.g., the use of CNNs in object classification, are a core topic of data science and are dealt with, for example, in the modules "Data Mining & Machine Learning" and "Artificial Intelligence". AI-supported image processing places high demands on the organization and processing of data at all levels of IoT processing concepts (edge/fog/cloud). This is the core of the module "Big Data Architectures". The consideration of the entire system process from the imaging sensor to the cloud requires a holistic view of the complete data science process, which is covered in the module "Data Science".

Available Resources

Ensuring the availability of data, computing resources, hardware, application experts

- Information required to create the scenario (system description, interfaces, documentation, relevant key figures, etc.) is provided by the resource person
- Steinel GmbH provides extensive test data sets from real production environments
- The Steinel GmbH contact person will be available for the duration of the project
- The components required for the prototype development as well as other required materials will be provided by the Steinel GmbH
- Hardware for the more complex machine learning is available via the Data Science Lab, the CfADS, and the AI computing cluster, yourAI, at Bielefeld University of Applied Sciences

Project Plan

First semester: Preparation of a research exposé as an examination. Familiarization with the concept of AI-based object detection algorithms, in particular, YOLO, R-CNN, image recognition, neural network approximation techniques, hardware platforms, and design flow tools.

Second semester: Creation of the system concept for design space exploration of cognitive edge computing architectures. Research on relevant work in the field of AI-based object detection algorithms and model compression in the above-mentioned context. Preparation of a paper that gives an overview of the respective research area, as an examination.

Third semester: Development of a first demonstrator and proof-of-concept for hardware acceleration for human presence detector (HPD) using, for example, YOLO-V7. Comparison of the developed framework with a classic implementation.

Fourth semester: Master's thesis and colloquium. Implementation and comparison of different combinations of AI-based object detection algorithms and hardware accelerators. Systematic evaluation and exploration of the efficiency of the developed framework. Final evaluation by comparing the implemented strategies with the state-of-the-art methods. Preparation of a paper with the first quantitative results as an examination.

Necessary Competencies

Mandatory:

- Expertise in Python, Pytorch
- Good knowledge of C++

Optional:

- Experience with hardware design flow tools
- Programming of microcontrollers/FPGAs
- Basic knowledge of HDL (Verilog, VHDL)
- Experience with IoT devices
- Experience with the version control system "git"

Acquirable Competencies

- Resource-efficient information processing at edge (embedded microcontrollers, FPGAs) in line with the IoT processing concept
- Sensor-related information processing
- AI/ML-based object detection methods
- Leveraging embedded hardware to accelerate AI/ML processes

Research Project for the Research Master Data Science

Project title: 8 Hardware-AI Codesign for Efficient Execution of Neural Networks

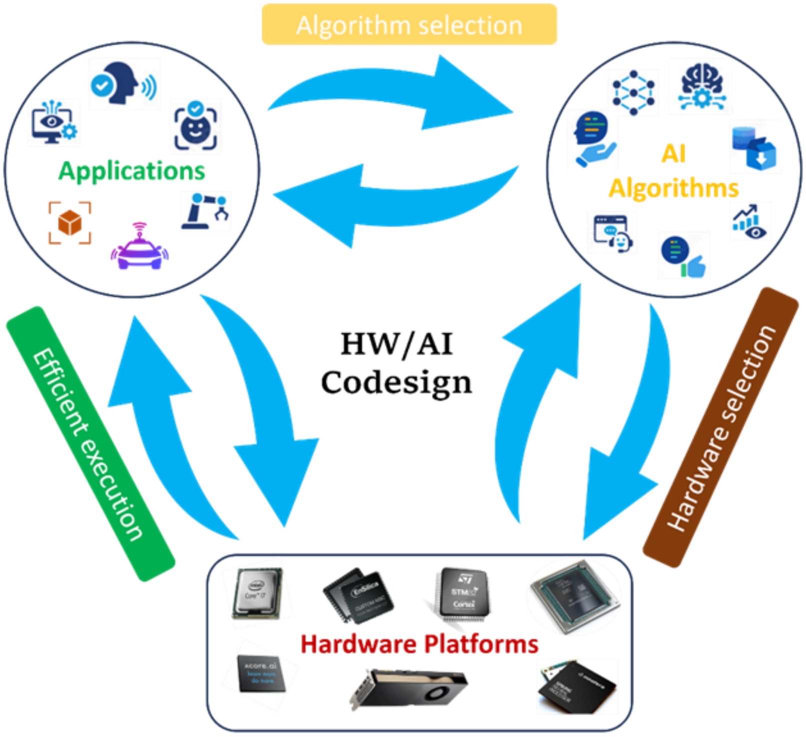


Figure 1: Design space exploration for Hardware-AI codesign

Project Overview

Number of Students	1
Project Type	Study project in collaboration with an external partner
Project Owner	Prof. Dr.-Ing. habil. Thorsten Jungeblut, Dr.-Ing. Qazi Arbab Ahmed
Project Context	**CareTech OWL** transfer project in collaboration with external partner, **Steinel GmbH**, as part of the CareTech OWL research project. Parallel employment as a student research assistant (WHK) is possible.

Abstract

The artificial intelligence (AI) paradigm has replaced a handful of traditional computer vision techniques to intelligently predict the correct output of a complex system under tight constraints and varying conditions. However, many AI algorithms require huge amounts of hardware resources due to their massive computational requirements. The goal of this project is the development of a framework

for the efficient implementation of AI and ML algorithms on highly resource-constrained hardware architectures. Based on the framework, automated methods for exploring the design space of suitable combinations of AI algorithms and hardware in terms of HW/AI-co-design will be explored. In particular, industrial applications with high latency requirements are targeted. Not only the entire chain as a linear process from model training to inference will be taken into account, but also the impact of the choice of possible hardware configurations on the original model development.

Short description

For the efficient execution of AI algorithms, techniques such as federated learning (FL) and cognitive edge computing (CEC) have already shifted the load of training and inference of neural networks from the cloud to the edge. The partitioning of the execution of the application has a significant impact on the performance and resource efficiency of the overall system. For example, a distinction can be made here between fundamentally different approaches of decentralized feature extraction as close as possible to the sensor with subsequent fusion versus central processing in the cloud. The first approach requires powerful edge hardware for preprocessing and potentially offers advantages for high realtime requirements. The second approach places higher demands on the communication infrastructure but potentially enables the execution of more complex networks. The main goal of this project is therefore to explore automated methods for exploring the design space of suitable combinations of AI methods and resource-constrained hardware in terms of HW/AI codesign.

In the area of efficient execution of AI procedures on embedded systems (cognitive edge computing), great progress has been made in the past. At all levels of the different processing concepts (cloud, fog, edge, very edge) in the production chain, a variety of potential hardware architectures and AI accelerators can be found, which differ in the available system resources (e.g., performance or power consumption). Examples of relevant hardware architectures are embedded microcontrollers with integrated AI acceleration, embedded GPUs/FPGAs, dedicated AI hardware accelerators, or high-end GPUs/FPGAs from the HPC domain.

We consider the complete processing chain, starting with the pre-processing close to the sensor (very edge), by connecting the groups of sensors and their information in the edge, via local decentralized cloud instances (fog), to the central station of all relevant information (cloud). At each level in the processing chain, individual approaches exist to locally optimize the resource efficiency of AI processes, for example, by reducing numerical precision (e.g., from 32 to 16, or 8 bits) to enable more efficient execution on specialized hardware or to minimize local storage space. The goal is to determine an optimal combination of hardware and software (i.e., AI algorithm) as part of a holistic design space exploration at all processing levels.

Task definition

In this project, students will develop an automated toolkit for design space exploration that can support the user in HW/AI codesign. In order to support the sustainable use of the developed methods and frameworks by the industrial partner, universally applicable models and automated design tools are to be made available in this project in the form of a development toolkit, which abstracts from the concrete use cases and make it possible to obtain recommendations for an optimal combination of AI methods and suitable (edge) hardware architecture based on a future application scenario. The goal is to provide the user with preliminary models and automated design tools that enable recommendations for a combination of AI methods and suitable (edge) hardware architecture based on a given application scenario and evaluation measures.

Figure 1 illustrates the general concept of design space exploration of hardware-AI codesign. The first main challenge is the compression/optimization of the neural network to reduce the size, yet with acceptable accuracy, using AI approximation techniques. The next challenge is to develop resource efficient techniques to map the compressed (AI) model to a suitable hardware platform, such as GPUs, CPUs, ASICs, and FPGAs, considering the resource utilization in terms of area, power/energy, and latency/throughput.

Reference to the topic of data science

The evaluation and application of AI/ML methods in the field of "machine vision", e.g., the use of CNNs in object classification, are a core topic of data science and are dealt with, for example, in the modules "Data Mining & Machine Learning" and "Artificial Intelligence". AI-supported image processing places high demands on the organization and processing of data at all levels of IoT processing concepts (edge/fog/cloud). This is the core of the module "Big Data Architectures". The consideration of the entire system process from the imaging sensor to the cloud requires a holistic view of the complete data science process, which is covered in the module "Data Science".

Available Resources

Ensuring the availability of data, computing resources, hardware, application experts

- Information required to create the scenario (system description, interfaces, documentation, relevant key figures, etc.) is provided by the resource person
- Steinel GmbH provides extensive test data sets from real production environments
- The Steinel GmbH contact person will be available for the duration of the project
- The components required for the prototype development as well as other required materials will be provided by the Steinel GmbH
- Hardware for the more complex machine learning is available via the Data Science Lab, the CfADS, and the AI computing cluster, yourAI, at Bielefeld University of Applied Sciences

Project Plan

First semester: Preparation of a research exposé as an examination. Familiarization with the concept of AI algorithms (CNN, DNN), in particular, object detection and image recognition, hardware platforms, and design flow tools used.

Second semester: Creation of the system concept for design space exploration of cognitive edge computing architectures. Research on relevant work in the field of the use of AI/ML methods for data processing and model compression in the above-mentioned context. Preparation of a paper that gives an overview of the respective research area, as an examination.

Third semester: Development of a first demonstrator and proof-of-concept for hardware acceleration of relevant application. Comparison of an AI process with a classic implementation.

Fourth semester: Master's thesis and colloquium. Implementation and comparison of different combinations of AI/ML methods and hardware accelerators. Systematic evaluation and exploration of the efficiency of the combinations. Comparison of different processing concepts (embedded AI, edge, cloud). Final evaluation by comparing the implemented strategies. Preparation of a paper with the first quantitative results as an examination.

Necessary Competencies

Mandatory:

- Expertise in Python, Pytorch
- Good knowledge of C++

Optional:

- Experience with hardware design flow tools
- Programming of microcontrollers/FPGAs
- Basic knowledge of HDL (Verilog, VHDL)
- Experience with IoT devices
- Experience with the version control system "git"

Acquirable Competencies

- Resource-efficient information processing at the various levels (edge, fog, cloud) in line with the IoT processing concept
- Sensor-related information processing
- AI/ML methods
- Leveraging embedded hardware to accelerate AI/ML processes

Research Project for the Research Master Data Science

Project title: 9 Cognitive Edge Computing for AI/ML-based Surface Inspection

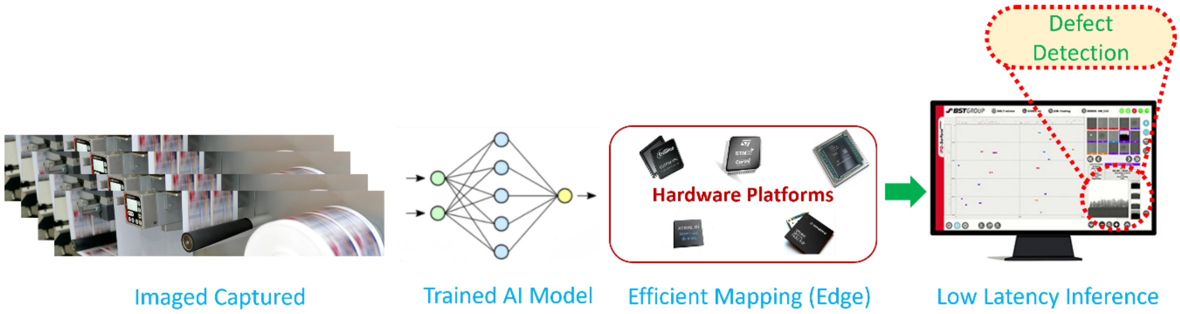


Figure 1: Efficient surface inspection using AI on edge

Project Overview

Number of Students	1
Project Type	Study project in collaboration with an external partner
Project Owner	Prof. Dr.-Ing. habil. Thorsten Jungeblut Dr.-Ing. Qazi Arbab Ahmed
Project Context	**CareTech OWL** transfer project in collaboration with an external partner, **BST Group**, as part of the CareTech OWL research project. The partner company **BST Group** provides extensive test data sets as well as a prototype hardware setup. Parallel employment as a student research assistant (WHK) is possible. The BST Group is very interested in the recruitment and long-term employment of junior staff.

Abstract

The artificial intelligence (AI) paradigm has replaced a handful of traditional computer vision techniques to intelligently predict the correct output of a complex system under tight constraints and varying conditions. The aim of the project is the design space exploration of AI/ML hardware accelerators in surface inspection. The focus is not only on learning the models on HPC systems but also on efficient execution (inference) on embedded hardware. The result of the design space exploration is the partitioning of the application, i.e. which AI methods can be evaluated on the sensor, which procedures can be accelerated via edge hardware (e.g. embedded GPU/FPGA), and which require powerful HPC hardware in the cloud.

Short description

Developments in the field of intelligent technical systems (ITS) are currently leading to a radical change in the entire value chain of industrial production. The increasing performance of information processing offers many new possibilities for the application area of industrial image processing, where humans and machines have so far reached their limits. In particular, the use of machine learning processes and artificial intelligence methods promises previously unknown possibilities, for example in object classification or visual quality control. Since around 2010, significant progress has been made with deep neural networks (DNN) and convolutional neural networks (CNN).

However, the performance of established AI/ML methods has so far mostly been based on the use of powerful decentralized computing resources (high-performance computing) in the cloud. The user is dependent on these powerful resources not only for learning the models but also for their execution (inference). However, the requirements arising in the field of industrial image processing deviate from the issues addressed by large providers of AI expertise due to high demands on low latency, real-time capability, or data locality. Issues such as maintainability, certifiability, or privacy also make it difficult to use popular models such as DNNs or CNNs in the cloud.

Nevertheless, great progress has also been made in the past in the area of efficient execution of AI/ML processes on embedded systems (**cognitive edge computing**). Suitable hardware accelerators can be found at all levels of the different processing concepts in networked production (edge/fog/cloud computing), which aim to achieve a suitable compromise between system resources such as required performance (e.g. classification accuracy), power consumption/energy requirements or data throughput/latency. Examples of relevant hardware architectures are embedded microcontrollers, embedded GPUs, embedded FPGAs, or dedicated AI hardware accelerators.

The BST Group develops systems for surface inspection (e.g., for battery cell production) that can reliably and immediately detect typical defects in the manufacturing process (e.g., coating discontinuities). BST systems can be perfectly adapted to a wide range of applications, due to their modular design. Image capture and defect detection can take place in real-time. The systems can be used on uniform, textured, and printed surfaces. The immediate and automatic detection and display of even the smallest defects and deviations enables the process to be adapted quickly and reliably to avoid rejects. The system consists of several optical sensors whose sensor data is suitably fused. The use of machine learning methods promises higher performance in defect detection and optimization of the entire process. However, with web speeds of several hundred meters per second, very high demands are required on the throughput and latency of the AI/ML processes used. Figure 1 shows the overview of the proposed methodology.

Task definition

The aim of the project is the design space exploration of AI/ML hardware accelerators for use in surface inspection. The focus is not only on learning the models of HPC systems but also on efficient execution (inference). Consideration of the entire system architecture, from intelligent sensors to edge gateways for local data pre-processing through to the cloud infrastructure, places high demands on the mapping of the AI network model to the hardware. The choice of target architecture, in turn, influences the choice and training of models. This results in an iterative cycle (model-to-inference-to-model), which includes both the selection of a suitable AI algorithm and the determination of the (hyper)parameters of the model. This approach of holistic design space exploration can also be referred to as HW/KI codesign, based on the established term HW/SW co-design. The design space exploration should consider numerous design goals, such as classification accuracy, latency, or resource requirements of the hardware. The result of the design space exploration is the partitioning of the application, i.e. which

AI algorithm can be evaluated directly on the sensor (e.g. through dimension reduction or feature extraction), which algorithm can be accelerated via edge hardware (e.g. embedded GPU/FPGA or dedicated TPU accelerators) and which require powerful HPC hardware in the cloud.

Reference to the topic of data science

The evaluation and application of AI/ML methods in the field of "machine vision", e.g., the use of CNNs in object classification, are a core topic of data science and are dealt with, for example, in the modules "Data Mining & Machine Learning" and "Artificial Intelligence". AI-supported image processing places high demands on the organization and processing of data at all levels of IoT processing concepts (edge/fog/cloud). This is the core of the module "Big Data Architectures". The consideration of the entire system process from the imaging sensor to the cloud requires a holistic view of the complete data science process, which is covered in the module "Data Science".

Available Resources

Ensuring the availability of data, computing resources, hardware, application experts

- Information required to create the scenario (system description, interfaces, documentation, relevant key figures, etc.) is provided by the BST Group
- The BST Group provides extensive test data sets from real production environments
- The BST Group contact person will be available for the duration of the project
- The components required for the prototype development as well as other required materials will be provided by the BST Group
- Hardware for the more complex machine learning is available via the Data Science Lab, the CfADS, and the AI computing cluster, yourAI, at Bielefeld University of Applied Sciences

Project Plan

First semester: Preparation of a research exposé as an examination. Familiarization with the concept of BST's surface inspection systems, and the interfaces of the intelligent sensors provided.

Second semester: Creation of the system concept for design space exploration of cognitive edge computing architectures. Research on relevant work in the field of the use of AI/ML algorithms for sensor data processing in the above-mentioned context. Preparation of a paper that gives an overview of the respective field of research as an examination achievement.

Third semester: Development of a first demonstrator and proof-of-concept for hardware acceleration of an application for surface inspection of BST. Comparison of an AI-based inspection with a classic implementation.

Fourth semester: Master's thesis and colloquium. Implementation and comparison of different combinations of AI/ML methods and hardware accelerators. Systematic evaluation and exploration of the efficiency of the combinations. Comparison of different processing concepts (embedded AI, edge, cloud). Final evaluation by comparing the implemented strategies. Preparation of a paper with the first quantitative results as an examination achievement.

Necessary Competencies

Mandatory:

- Expertise in Python, Pytorch
- Good knowledge of C++

Optional:

- Basic knowledge of HDL (Verilog, VHDL)
- Experience with hardware design flow tools
- Programming of microcontrollers/FPGAs
- Experience with IoT devices

Acquirable Competencies

- Resource-efficient information processing at edge (embedded microcontrollers, FPGAs) in line with the IoT processing concept
- Sensor-related information processing
- AI/ML-based object detection methods
- Leveraging embedded hardware to accelerate AI/ML processes

Research Project for the Research Master Data Science

Project title: 10 Optical quality control of coated wood blanks

Project overview

Number of Students	1
Project Type	Project with external partners
Project Owner	Prof. Dr.-Ing. Christian Schwede
Project Context	Project in cooperation with the company **Febrü Büromöbel GmbH** and the Institute for Data Science Solutions (IDaS); **employment as a student assistant** by the university is planned.

Abstract

The aim of the project is to visually inspect the quality of coated wood veneers that are cut and edged in a wide variety of office furniture production processes. Both the basic setup (camera procurement and suspension, data collection and real-time data processing) and the process based on deep convolutional neural networks are to be developed.

Short description

Febrü is a medium-sized company based in Herford that specializes in the production of office furniture. The company employs around 250 people and generates an annual turnover of around 45 million euros. Febrü produces exclusively in Germany and focuses on sustainable production with high quality standards and individual solutions for the modern working world. After the wood decors have been cut to size, a manual quality check is currently carried out by an employee. This employee checks the dimensions (length, height and width) for compliance with the specifications and the surface for damage. The surfaces have a wide range of variants and differ not only in the printed pattern (wood types, concrete look, etc.) but also in the different reliefs. In future, the task is to be supported by camera-based methods from the field of machine learning.

Task definition

In the project, the student should firstly set up the setup required to record the data and carry out the optical quality inspection. This includes selecting the camera, installing it in production, setting up the recording and coordinating with the quality inspection staff so that they can digitally and systematically record the results of their inspection. Secondly, methods from the field of image analysis and machine learning are to be selected and implemented to enable optical recognition of dimensions and damage.

Reference to the topic of data science

The methods of machine learning and convolutional neural networks used are part of the Research Master's course and core areas of data science.

Available Resources

- Technical experts from production and quality management are provided by Febrü
- The hardware and software required for data collection and processing in production is procured by Febrü
- Data labeling is carried out by the Febrü QM expert
- Hardware for machine learning is available via the Data Science Lab, CfADS and the AI computing cluster yourAI at HSBI

Project plan

First semester: Preparation of a research exposé as examination performance. Familiarization with the task in the company, research on available camera hardware and software, procurement of hardware and software, setting up the setup in production, capturing the images and starting the labelling process.

Second semester: Literature research on optical image recognition methods and convolutional neural networks (CNN) in this context. Implementation of methods for the optical determination of dimensions. Preparation of a paper that gives an overview of the research area is an examination achievement.

Third semester: Implementation and training of CNN with the collected data. Evaluation of the results. Publication of a paper with the first results is an examination achievement.

Fourth semester: Master's thesis and colloquium. Improvement of the results and integration of the algorithm into the production environment.

Necessary Competencies

Mandatory:

- Programming skills (especially Python)
- Knowledge of real-time programming and sensor data processing (automation/mechatronics)

Optional:

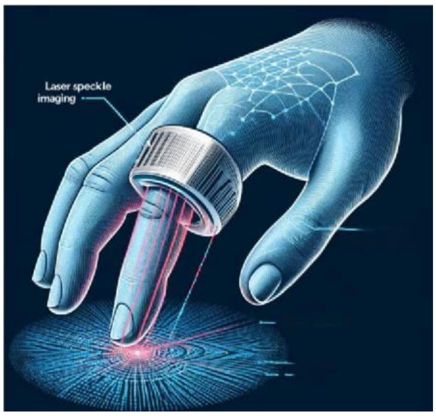
- Experience with image processing algorithms
- Experience with machine learning
- Experience in production and quality management

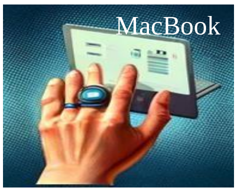
Acquirable Competencies

- Image processing
- Optical quality control
- Convolutional Neural Networks
- Real-time data processing
- Publication of scientific papers at international conferences
- Skills in working in a team

Research Project for the Research Master Data Science

Project title: 11 TextureSense: Enhancing Object Recognition and Activity Awareness with Wearable Laser Speckle Imaging





Texture detection by touching the surface of an object for activity awareness.

Project overview

Number of Students	1
Project Type	Study project
Project Owner	Prof. Dr.-Ing. habil. Thorsten Jungeblut Dr. Hasina Attaullah
Project Context	TransCareTech project Parallel employment as a student research assistant (WHK) is possible

Abstract

Continuous monitoring of human activities through real-time technologies has significant prospective in numerous sectors, including healthcare, assisted living, and human-computer interaction (HSI). Wearable sensors that collect and analyse physiological signals are particularly favourable for this objective. This proposal focuses on utilizing wearable laser speckle imaging (LSI) technology to detect texture, enhancing awareness of everyday human activities. LSI technology provides non-invasive, continuous tracking of subtle texture variations, offering a new method for real-time activity recognition. This research intends to enhance the boundaries of activity monitoring through wearable LSI devices, leading to improved healthcare interventions, personalized daily assistance, and better human-computer interaction. The study will integrate data science, biomedical engineering, and human-computer interaction, utilizing machine learning to analyse texture and recognize activities from LSI data. This interdisciplinary approach is set to transform activity monitoring and promote innovation in personalized healthcare and assistive technology.

Short Description

LSI is a non-invasive optical technique that measures the dynamic speckle pattern formed by coherent light scattering from a surface. It has been widely used in biomedical applications for blood flow imaging, tissue perfusion monitoring, and skin assessment. Recently, there has been growing interest in applying LSI to capture surface texture changes associated with human activities. By analysing the temporal variations in speckle patterns, it is possible to infer different textures such as book, mug, face, mouse and orange etc. and then identify the human activities.

Wearable technology has revolutionized real-time texture prediction. Using wearable LSI, this research intends to develop a system for human activity awareness by detecting objects based on texture changes, suggesting innovations in personalized healthcare and human-computer interactions.

Initial Situation, Motivation, and Objectives

Current activity monitoring technologies often rely on direct physical or visual sensors, which can be invasive or limited in scope. Wearable LSI offers a non-invasive, comprehensive method for detecting subtle physiological changes. The objective is to develop an LSI-based wearable that can interpret these changes as routine human activities, aiming to improve interventions in healthcare and enhance interaction mechanisms in HCI.

The primary objective of this research is to develop activity awareness based on texture detection systems using wearable laser speckle imaging for object detection. Specifically, we aim to:

- Investigate the feasibility of using LSI to capture and analyse surface texture changes associated with various textures.
- Develop signal processing algorithms to extract relevant features from LSI data for texture classification.
- Implement a wearable LSI device prototype capable of real-time texture detection and activity inference.
- Evaluate the performance of the proposed system through experiments involving human subjects performing daily activities.

Task Definition

The project will focus on:

1. Collect and process LSI data.
2. Developing algorithms for feature extraction from LSI data.
3. Designing machine learning models for texture classification for activity awareness.
4. Building a prototype of a wearable LSI device.
5. Conducting experiments to validate the system's efficacy.

Concrete Task of the Students

Students will engage in:

1. Data Collection: Acquire LSI data from a diverse set of routine activities performed by human subjects in a controlled laboratory setting.
2. Feature Extraction: Develop algorithms to extract texture-related features from LSI data, such as spatial frequency, contrast, and temporal dynamics of speckle patterns.
3. Activity Classification: Employ machine learning techniques, such as support vector machines (SVMs) or deep neural networks (DNNs), to classify textures based on extracted features.
4. Wearable Device Development: Design and build a wearable LSI device prototype that can capture and process speckle imaging data in real-time.

5. Evaluation: Validate the performance of the proposed system through experiments involving human subjects. Assess the accuracy of activity classification and the robustness of the wearable device in real-world scenarios.

Reference to the Topic of Data Science

This research integrates data science with biomedical engineering and HCI through:

- Feature extraction and pattern recognition from physiological signals.
- Employing machine learning for real-time activity classification.
- Utilizing statistical methods for robust data analysis.

Available Resources

Resources will include but not available:

- Access to LSI devices and wearable sensor components.
- Software for data analysis and machine learning.
- Collaboration with experts in biomedical engineering and HCI.

Project Plan

First Semester:

1. Participate in workshops/seminars on recent trends and technologies in human computer interaction.
2. Conduct a literature review to understand current technologies and methodologies in activity monitoring and texture detection.
3. Submit a literature review report.
4. Collect real time LSI data.
5. Present a project proposal outlining the scope and goals for the upcoming semesters.

Second Semester:

1. Writing a paper of a literature review is the mandatory artefact.
2. Design and develop the initial prototype using wearable LSI technology.
3. Run initial test to assess the prototype's functionality in detecting texture changes.
4. Test results report and analysis.
5. Revised project plan based on feedback and test results.

Third Semester:

1. Refine the prototype based on feedback and detailed analysis from the second semester.
2. Develop and integrate machine learning model for better accuracy in activity awareness.
3. Conduct comprehensive testing** both in controlled environments and real-world settings.
4. Write a paper with the first results is the mandatory artefact.

Fourth Semester:

1. Final refinements and testing of the system to ensure reliability and accuracy.

2. Prepare a detailed project report that includes all aspects of research, development, testing, and conclusions.
3. Explore possibilities for publication in academic journals or presentations at conferences.
4. Master thesis and colloquium

Each semester should also include regular meetings with supervisors or mentors, maintaining a project log, and scheduling reviews to track progress and address any issues. Adjustments may be necessary based on the project's evolution, feedback from advisors, and practical considerations.

Necessary Competencies

Mandatory:

- Proficiency in data science and machine learning.
- Basic knowledge of signal processing and physiological data interpretation.

Optional:

- Experience with wearable technology development.

Acquirable Competencies

Students will acquire:

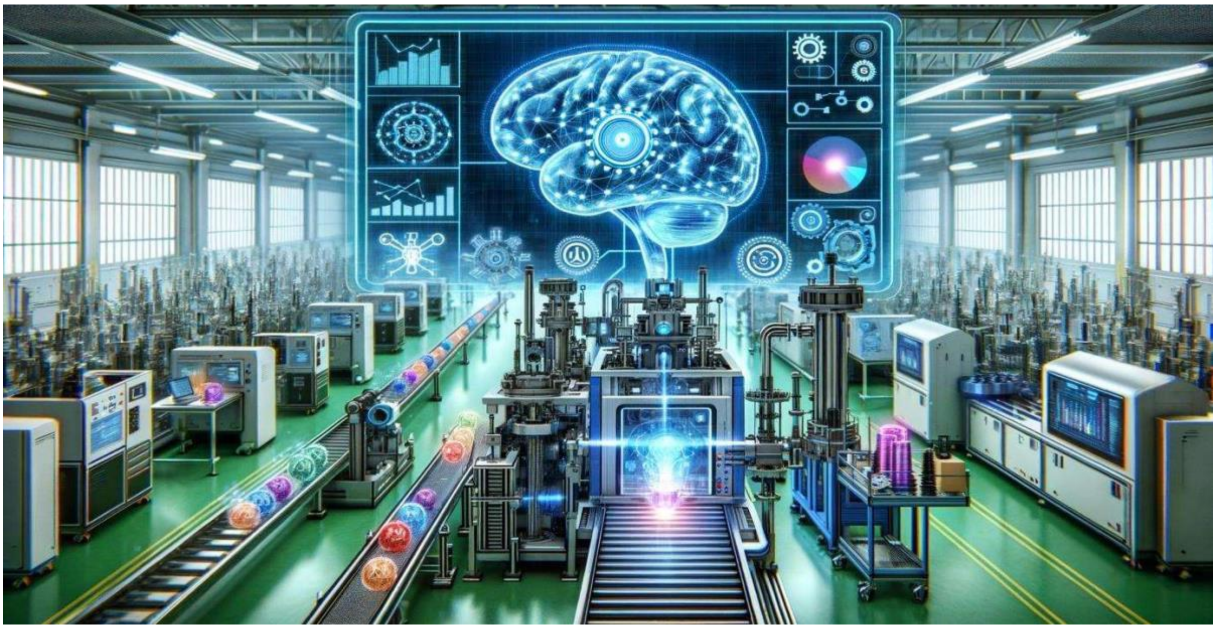
- Advanced skills in data science applied to physiological data.
- Experience in interdisciplinary research and prototype development.
- Expertise in handling and analysing real-time data from wearable devices.

Reference:

Wang, Xue, and Yang Zhang. "TextureSight: Texture Detection for Routine Activity Awareness with Wearable Laser Speckle Imaging." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7.4 (2024): 1-27.

Research Project for the Research Master Data Science

Project title: 12 **AI supported Production Planning for Mass Customization (KiMaC)**



Project overview

Number of Students	1-2
Project Type	Funded project with partner companies
Project Owner	Prof. Dr. Pascal Reusch
Project Context	An interdisciplinary research initiative in collaboration with industry leaders focused on the implementation of Mass Customization using AI techniques. Partners include leading manufacturing companies. Opportunities for employment as a research assistant may be available.

Abstract
This project aims to revolutionize production through Mass Customization, combining the **personalization** of custom-made products with the **efficiency and cost-effectiveness of mass production**. The project leverages AI technologies to **tackle real-world challenges**, focusing on adapting production processes for customizable product offerings and improving efficiency. This includes **AI-assisted systems** to streamline configuration, production logistics, and control processes, significantly reducing lead times and errors while enhancing precision and lowering costs.

Short description
The KiMaC project represents a pioneering initiative in the field of mass customization, combining the strengths of **GRAUTHOFF Türengruppe GmbH** and **Solarlux GmbH** with cutting-edge artificial intelligence to revolutionize the production of doors and glass folding walls. This initiative not only emphasizes the technological innovation but also the **sustainability and customization capabilities** that AI brings to the manufacturing processes. GRAUTHOFF Türengruppe GmbH is a renowned player in the building industry, specializing in the manufacturing of doors. With over 60 years of history that began in a small workshop, GRAUTHOFF has grown into one of Germany's most significant and advanced door manufacturers. Solarlux GmbH has established itself as a leader in the production of glass folding walls, conservatories, and facade solutions for over 35 years. Solarlux is celebrated for its exceptional quality, sustainable practices, and innovative designs.

The KiMaC project aims to leverage AI to streamline the customization process, allowing for the efficient production of tailor-made products that **meet individual customer preferences without sacrificing the efficiency of mass production**. By integrating AI into the product configuration and manufacturing logistics, KiMaC facilitates a new level of personalization in the production of doors and glass folding walls, ensuring that each product is not only functionally superior but also uniquely suited to the needs and tastes of its users.

On one hand, product configuration will be restructured with a knowledge graph as its new database, while on the other, **production planning will be adapted** to enable the manufacturing of customized products at mass production costs. This part of the project focuses on the production planning part **applying AI supported solutions in the real world**. The KiMaC project not only represents a step forward in manufacturing efficiency and product personalization but also highlights the participating companies' commitment to innovation, sustainability, and customer satisfaction in the modern era.

Task definition
Students will explore and contribute to various aspects of the project, including:

- Data (pre-)processing & data analytics of real-world company data.
- Creation of AI-based assistant systems for the production planning and configuration process, offering production plans and decision support
- Automation of manufacturing order scheduling and workforce planning using AI, aligning with real-time operational data and current order statuses.
- Integration and evaluation of assistant systems within partner companies' production systems, assessing economic viability and acceptance.

Reference to the topic of data science
The project encapsulates core Data Science methodologies, including data pre-processing, data analytics and machine learning especially reinforcement learning, **applied to real-world production**

optimization challenges. It exemplifies the practical application of theoretical knowledge in industry settings, offering a unique learning and development opportunity in the intersection of AI and manufacturing.

Available Resources

Collaboration with industry partners provides access to real production environments, operational data, and IT infrastructures for the development and implementation of AI models and systems. Additionally, the **Centre for Applied Data Science Gütersloh** offers access to a **cutting-edge test IoT-Factory** for further hands-on experience and testing. A modern cloud environment will be utilized for software deployment, supported by extensive computational resources for ML training and optimization tasks provided by the **academic institution's Data Analytics Cluster**.

Project plan

First Semester: Research exposé detailing initial AI model conceptualization for production optimization.

Second Semester: Systematic literature review on existing AI applications in mass customization and production optimization.

Third Semester: Development and initial testing of AI models and assistant systems within a simulated production environment.

Fourth Semester: Master thesis focusing on the integration, evaluation, and optimization of AI systems in real-world production settings, followed by a colloquium.

Necessary Competencies

Mandatory:

- Bachelor's degree with a focus on data science, production logistics, computer science, or a related field.
- Programming skills, preferably Python.
- Basic understanding of machine learning, statistics, and optimization.
- Interest in applying data science to real industrial challenges.

Optional:

- Experience with Operations Research, Reinforcement Learning, or process simulation.
- Knowledge in production planning, ERP systems, or related domains.

Acquirable Competencies

- **Practical application** of advanced machine learning and AI in complex, **real-world production settings**.
- Deep understanding of **mass customization challenges and solutions** from a data science perspective.
- **Enhanced research skills** through the development, implementation, and evaluation of cutting-edge AI systems in an interdisciplinary setting.

Research Project for the Research Master Data Science

Project title: 13 Digital twin for batch size optimization of a powder coating system

Project overview

Number of Students	1
Project Type	Project with external partners
Project Owner	Prof. Dr.-Ing. Christian Schwede
Project Context	Project in cooperation with the company **Febrü Büromöbel GmbH** and the Institute for Data Science Solutions (IDaS); **employment as a student assistant** by the university is planned.

Abstract

The aim of the project is to optimize the batch sizing of a powder coating system for metal parts of a multi-variant office furniture production. For this purpose, the production and storage processes around the plant as well as from the company's data must be automatically learned as a digital twin and an AI assistant for batch size planning must be developed on this basis.

Short description

Febrü is a medium-sized company based in Herford that specializes in the production of office furniture. The company employs around 250 people and generates an annual turnover of around 45 million euros. Febrü produces exclusively in Germany and focuses on sustainable production with high quality standards and individual solutions for the modern working world. An expensive production step in the manufacture of metal parts is the powder coating of the metal parts to give them different colours. The metal parts are highly varied and are produced from sheets or tubes by laser cutting, punching, welding and grinding. In the powder coating plant, they have to be combined into colour lots so that as few colour changes as possible have to be carried out (set-up time reduction) and delivery reliability is maintained at the same time. Another criterion is that the colours are applied from light to dark. One difficulty here is that the metal parts are often not provided as planned from the previous production steps, meaning that the planning has to be adapted flexibly. A few experienced employees are currently responsible for planning.

Task definition

In the project, the student should first create a digital twin from production and logistics data using the open source framework OFacT. Machine learning and process mining methods will be used to create the simulation-based digital twin as automatically as possible. Secondly, an AI assistant for batch size planning is to be created on the basis of the digital twin. Mathematical optimization and reinforcement learning methods can be used here. The knowledge of the human planners is to be integrated through an interaction and feedback mechanism to be developed (possibly based on large language models).

Reference to the topic of data science

The methods of machine learning, reinforcement learning and large language models used are part of the course in the Research Master and core areas of data science.

Available Resources

- Data and technical experts from production are provided by Febrü
- The OFacT framework is provided by IDaS
- Hardware for machine learning is available via the Data Science Lab, CfADS and HSBI's AI computing cluster yourAI

Project plan

First semester: Preparation of a research exposé as an examination. Familiarization with the task in the company, inspection of the data, familiarization with the OFacT framework. First manual modelling of the digital twin

Second semester: Literature review on optimization and reinforcement learning methods for production planning. Automatic creation of the digital twin from the data using machine learning and process mining. Preparation of a paper that provides an overview of the research area is an examination achievement.

Third semester: Implementation of the AI assistant and evaluation of the results. Publication of a paper with the first results is an examination achievement.

Fourth semester: Master's thesis and colloquium. Implementation of the procedure for integrating human feedback.

Necessary Competencies

Mandatory:

- Programming skills (e.g. Python)

Optional:

- Experience with machine learning
- Experience with optimization processes
- Experience in production and logistics
- Experience with simulation models

Acquirable Competencies

- Digital twins
- Reinforcement Learning
- Large Language Models
- Optimization methods for production planning
- Production and logistics know-how in variant-rich discrete manufacturing
- Publication of scientific papers at international conferences
- Skills in working in a team

Research Project for the Research Master Data Science

Project title: 14 Optimization of the panel warehouse of a medium-sized office furniture manufacturer

Project overview

Number of Students	1
Project Type	Project with external partners
Project Owner	Prof. Dr.-Ing. Christian Schwede
Project Context	Project in cooperation with the company **Febrü Büromöbel GmbH** and the Institute for Data Science Solutions (IDaS); **employment as a student assistant** by the university is planned.

Abstract

The aim of the project is to optimize the storage of wooden panels for office furniture production. The difficulty lies in the fact that the production orders are not yet fully known when the boards are stored in different stacks and the requirements must therefore be predicted on the basis of historical data. Machine learning models are to be used for this purpose.

Short description

Febrü is a medium-sized company based in Herford that specializes in the production of office furniture. The company employs around 250 people and generates an annual turnover of around 45 million euros. Febrü produces exclusively in Germany and focuses on sustainable production with high quality standards and individual solutions for the modern working world. An important production step in the varied production is based on externally supplied chipboard in around 90 different decors. In addition to a small external warehouse, a production warehouse with 64 storage spaces is available for this purpose, where the boards are stored in stacks. Individual access to the boards is therefore not possible and the lower boards can only be removed with time-consuming relocation. To make access as efficient as possible, the panels must be stored in such a way that the parts that are needed first in production are at the top of the stacks. The difficulty, however, is that not all production orders are available at the time of storage.

Task definition

In the project, the student should use methods from the field of machine learning to forecast a case-specific realistic load scenario for production at the time of put away based on historical and currently available orders. Based on this load scenario, an optimization algorithm is then to be implemented that calculates the best possible put away sequence and stacking location allocation. The solution is to be validated with real production data.

Reference to the topic of data science

The machine learning methods used are part of the course in the Research Master and core areas of data science.

Available Resources

- Data and technical experts from production are provided by Febrü
- Hardware for machine learning is available via the Data Science Lab, CfADS and HSBI's AI computing cluster yourAI

Project plan

First semester: Preparation of a research exposé as an examination. Familiarization with the task in the company, review of the data and definition of the optimization criteria.

Second semester: Literature review on machine learning methods, in particular on order forecasting. Implementation and an initial procedure. Preparation of a paper that provides an overview of the research area is an examination requirement.

Third semester: Comparison of different methods for order forecasting and evaluation of the results. Publication of a paper with the first results is an examination achievement.

Fourth semester: Master's thesis and colloquium. Implementation of the optimization method based on the predicted orders and validation of the results with real production data.

Necessary Competencies

Mandatory:

- Programming skills (e.g. Python)

Optional:

- Experience with machine learning

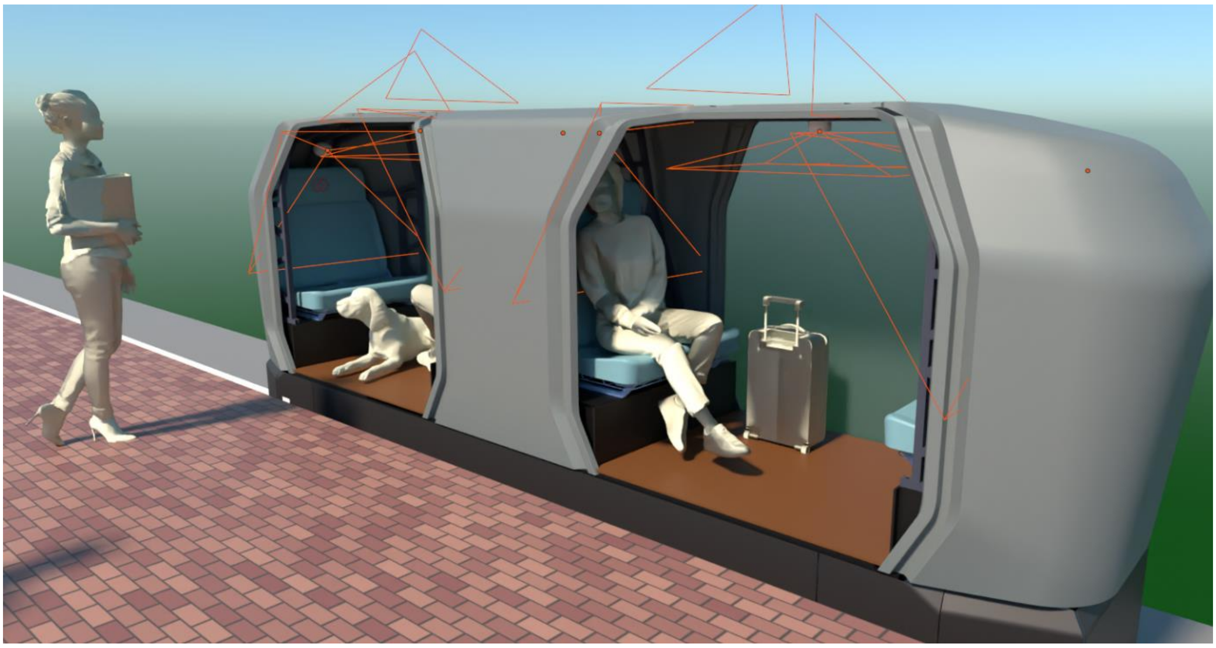
- Experience with optimization methods
- Experience in production and logistics

Acquirable Competencies

- Machine learning methods for order forecasting
- Optimization methods for production control
- Production and logistics know-how in variant-rich discrete manufacturing
- Publication of scientific papers at international conferences
- Skills in working in a team

Research Project for the Research Master Data Science

Project title: 15 Sensor Fusion and AI-Based Data Analysis for Interior Monitoring in Autonomous Trains



Project overview

Number of Students	1
Project Type	Funded research project
Project Owner	Prof. Dr. Thorsten Jungeblut
Project Context	Collaboration with the **enableATO** project; parallel employment as a research assistant is possible

Abstract

The aim of the project is to develop a system for monitoring the interior of autonomous trains using sensor fusion and AI methods. Different sensor types are combined in order to reliably recognize dangerous situations. The scientific challenge lies in the selection of suitable sensors, the evaluation of different fusion methods and the application of AI algorithms. The end result is a solution that ensures robust monitoring and can therefore improve safety in autonomous trains.

Short description

The mobility of the future must be resource-efficient, user-orientated and innovative. The **enableATO** project addresses these requirements by developing autonomous rail vehicles (Grade of Automation Level 4) for the use in rural areas. The aim is to use smaller, automated vehicles to promote sustainable and efficient mobility and to enable the reactivation of disused lines.

A key aspect of autonomous trains without **attendants** is ensuring safety and user-friendliness. The interior of these trains must be continuously monitored in order to recognize events such as falls, the need for assistance or unforeseen emergencies at an early stage. Different sensor types such as RGB cameras, radar and infrared each offer specific advantages, but reach their limits under adverse conditions or in complex scenarios.

Sensor fusion combines different sensor types to improve the robustness and accuracy of the monitoring systems. In addition, analysing sensor data using **machine learning and deep learning methods** plays a crucial role in managing complex indoor scenarios. AI-based approaches enable the precise analysis of movement patterns, the detection of dangerous situations and the implementation of real-time reactions, for example in the event of falls or abnormal passenger behaviour.

The project investigates how sensor fusion and AI-based data analysis can be used for interior monitoring in autonomous trains in order to minimize safety risks and maximize user-friendliness. As part of **enableATO**, technological foundations are being created for the mobility of tomorrow that will make rail transport systems more efficient, intelligent and safe.

Task definition

In this project, the students are developing a system for sensor fusion and AI- based interior monitoring that is specifically designed to detect dangerous situations in autonomous trains.

For this purpose, suitable sensor types are to be identified that are adequate for the specific application and complement each other in their functionality. Furthermore, ML and AI algorithms as well as fusion methods will be selected and evaluated in order to effectively combine the sensor data and to ensure precise hazard detection. The investigation includes the training and comparison of algorithms for the selected sensor combinations.

The aim is to develop a prototype that combines a suitable set of sensors and utilizes appropriate AI models and fusion methods. The work should determine which sensors, AI algorithms and fusion techniques offer the best performance for the use case.

Reference to the topic of data science

The evaluation and application of AI/ML methods, such as the use of convolutional neural networks (CNNs) in image recognition or time series analyses, are key topics in data science and are covered in depth in the modules “Advanced Machine Learning” and “Artificial Intelligence”. In addition, the focus is on the collection of data by various sensors, which places high demands on the organization and processing of data and is covered in the module “Big Data Architectures”. Last but not least, data analysis is an essential part of the project, which is covered in detail in the module “Data Mining & Machine Learning”.

Available Resources

- **System information:** Information required to create the scenario (e.g. system description, relevant key figures) is provided by the contact person.
- **Sensors:** Required materials such as the sensors to be investigated are provided by the enableATO project
- **Mockup of the vehicle interio**r: A mockup of the vehicle interior is available in which test data can be recorded.
- **Computing resources:** Hardware for the more complex machine learning is available via the Data Science Lab, CfADS and the AI computing cluster yourAI at HSBI.
- **Project support:** A contact person from enableATO will be available for the duration of the project.

Project plan

First Semester: Preparation of a research exposé is an examination task. Familiarization with the concept of various sensor technologies, fusion methods and AI algorithms (CNN, DNN), in particular for object recognition and image recognition.

Second Semester: Creation of a system concept for sensor fusion with a focus on the specific application. Research on relevant work in the field of sensor fusion for indoor surveillance and the corresponding use of AI methods. Preparation of a paper that provides an overview of the respective field of research area is an examination task.

Third Semester: Practical implementation of sensor fusion with specific sensors and own data collection. Implementation of various fusion methods and machine learning processes for danger detection and their evaluation. Preparation of a paper with initial quantitative results as an examination task.

Fourth Semester: Master's thesis and colloquium. Final evaluation by comparing the implemented strategies.

Necessary Competencies

Mandatory:

- Good experience with Python

Optional:

- Experience with the version control system “git”
- Basic experience with data analyses
- Basic experience with Python libraries such as PyTorch, Tensorflow or Scikit-learn

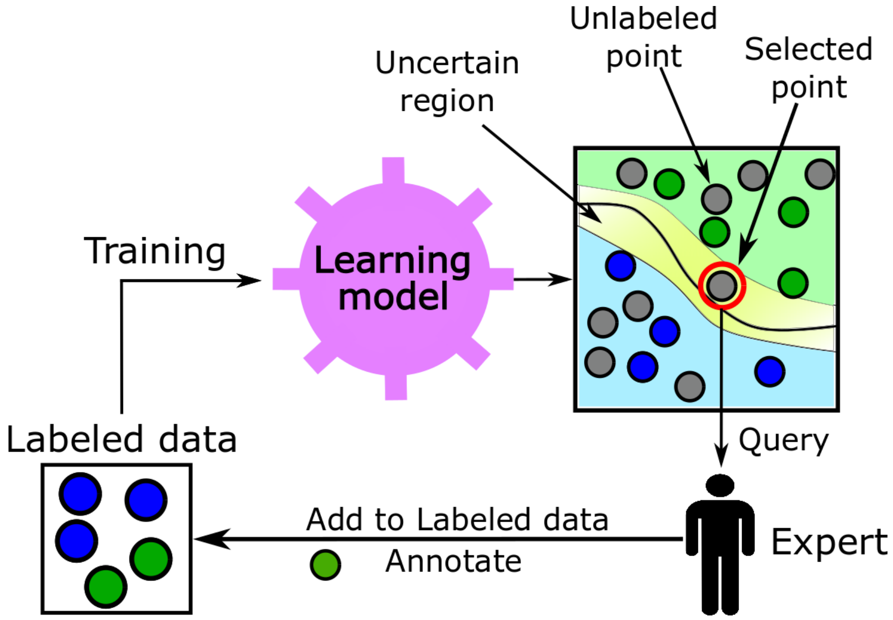
Acquirable Competencies

Competencies that can be acquired by the project.

- Sensor fusion
- AI/ML methods
- Sensor-based data processing
- Expertise in working in a team

Research Project for the Research Master Data Science

Project title: 16 Uncertainty Quantification in Active Learning



Project overview

Number of Students	1-2
Project Type	Master Student Project
Project Owner	Dr. Peter Kuchling Dr. Alaa Othman Prof. Dr.-Ing. Wolfram Schenck
Project Context	Project within the Center for Applied Data Science Gütersloh (CfADS) with internal university partners.

Abstract

Despite the vast amounts of data generated by IoT factories, healthcare systems and various industries, much of it remains 'unlabelled', limiting its usefulness for machine learning (ML) models. Labelling data can be expensive and time-consuming, creating a barrier to effective model development. Active Learning (AL) addresses this issue by enabling models to selectively query the most informative and representative data points. This research proposal investigates "Uncertainty Quantification in Active Learning", focusing on improving the selection process within AL frameworks. Using methods such as Bayesian inference and Gaussian processes, together with ensemble techniques, we will investigate how uncertainty quantification affects the efficiency of active learning. The study aims to improve the decision-making process in data selection, potentially advancing applications in areas where obtaining labelled data is difficult and costly.

Short description

In the rapidly evolving machine learning landscape, raw or unlabelled data is becoming very large and free in many areas such as IoT, healthcare and industry. Despite the massive amount of data being generated, the cost and time associated with labelling this data is hindering effective model training. The key challenge is to identify which data points will provide the most value when labelled, ensuring efficient learning without overwhelming resources. The active learning (AL) technique offers a solution by querying the most informative and representative points to form a small but high-quality training data that improves the performance and generalization of learning models. However, there are many AL strategies. In this proposal, we propose to incorporate uncertainty quantification into AL algorithms. By using methods such as Bayesian inference and Gaussian processes, our approach enhances the ability of the model to assess its confidence in predictions. This allows intelligent selection of the most informative data points, rather than random sampling. As a result, we can optimise labelling efforts, reduce costs and improve model performance.

In practice, this methodology is particularly beneficial in environments where data is abundant, but labelling is expensive. For example, in a healthcare setting, active learning can identify uncertainties in patient data and guide targeted labelling efforts to improve diagnostic accuracy. This strategic focus not only streamlines the learning process, but also maximizes the use of limited resources, paving the way for more effective machine learning applications across multiple domains.

Task definition

The student(s) will develop and implement a framework for uncertainty quantification in active learning algorithms. This framework will focus on enhancing the selection process for labelling data points in scenarios where unlabelled data is abundant. The goal is to integrate techniques such as Bayesian inference and Gaussian processes to accurately assess uncertainty in model predictions, thereby identifying the most informative data points for labelling. Students will work on refining the algorithm to ensure that it intelligently queries data based on uncertainty metrics, optimizing the labelling process and reducing costs associated with data annotation. The practical application of this framework will be demonstrated in a relevant industrial or healthcare setting, where the effectiveness of active learning can be evaluated in real-world scenarios. By focusing on uncertainty quantification, this project aims to improve model performance and efficiency, ultimately facilitating more effective machine learning applications in data-scarce environments.

Reference to the topic of data science

The proposed research aligns closely with fundamental data science principles by integrating uncertainty quantification into active learning frameworks. This approach complements the research master's curriculum, which emphasizes advanced methodologies in predictive modelling, data analysis, and decision-making based on data insights. The project offers students the opportunity to explore and apply innovative data science techniques in a practical setting, enhancing their understanding of how uncertainty quantification can improve the efficiency and effectiveness of machine learning models. By focusing on the strategic selection of data points for labelling, students will gain valuable experience in optimizing data-driven processes, ultimately contributing to more effective applications in various fields, including IoT and healthcare.

Available Resources

- **Expert Supervision**: The project benefits from the support of knowledgeable and experienced supervisors who are proficient in data science techniques, including active learning, as well as IoT and industrial applications. Students can draw on this expertise for guidance, feedback, and mentorship during the research process, helping to ensure the quality and relevance of their work.
- **Collaborative spaces**: The CfADS group at HS Bielefeld offers collaborative spaces for students to connect with each other, exchange ideas, and access extra resources. These hubs create an atmosphere that supports interdisciplinary cooperation and encourages a comprehensive approach to solving problems.
- **No additional hardware required**: The project utilizes the current hardware resources in the IoT factory, avoiding the necessity for further investment. Students can easily incorporate their research into the existing infrastructure, facilitating the implementation of adaptive sensor activation through active learning algorithms.

Project plan

First Semester: Project Setup and Exploration

- Conduct a literature review on uncertainty quantification and active learning applications.
- Identify challenges and gaps in integrating uncertainty quantification into active learning.
- Familiarize with theoretical foundations of uncertainty estimation techniques.
- Write a literature review paper summarizing your findings.

Second Semester: Initial Development and Prototyping

- Design the architecture for the active learning framework.
- Implement the framework using appropriate programming languages.
- Integrate active learning algorithms for data selection.
- Conduct initial experiments to evaluate performance.
- Analyse results to identify improvements.

Third Semester: System Refinement and Integration

- Evaluate the model on relevant datasets.
- Compare results with baseline and state-of-the-art methods.
- Identify limitations and refine the framework based on findings.
- Write a research paper detailing the evaluation and improvements.

Fourth Semester: Optimization and Final Evaluation

- Explore various experimental scenarios, including dataset sizes.
- Analyse the framework's impact on model performance, convergence speed, and labelling effort.
- Investigate scalability on large datasets and specific applications.
- Summarize results, draw conclusions, and recommend future research directions.
- Write the Master Thesis documenting the entire research process and findings.

Necessary Competencies

Mandatory: Strong programming skills, particularly in languages suitable for ML applications (e.g., Python).

Optional:

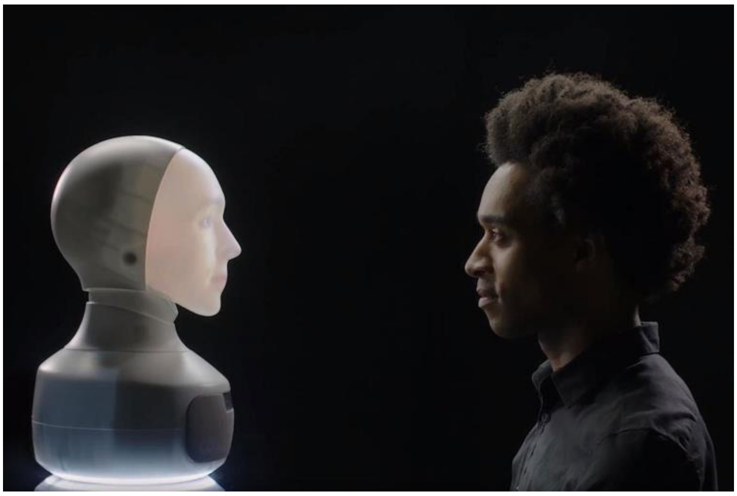
- Good mathematical background.

Acquirable Competencies

- Proficiency in uncertainty quantification for active learning.
- Experience with machine learning libraries and programming languages.
- Skills in evaluating model performance against benchmarks.
- Ability to conduct literature reviews and write research papers.
- Experience in designing experiments to assess model variables.

Research project for the Research Master Data Science

Project title: 17 Human-robot interaction: design, implementation and use of a humanoid robot as an individual conversation partner



Project overview

Number of students	1 to 2
Kind	Study project
Project responsibility	Prof. Dr. Thomas Süße and Dr. Maria Kobert
Project context	Possible employment as a student assistant depends on available financial resources. Cooperation with companies is also conceivable.

Abstract

The interaction between humans and social humanoid robots is central in current and future human-machine interaction. Social robots are able to communicate with people in a human-like way - by speaking, listening, showing emotions and maintaining eye contact. For example, they can be used to serve customers, provide companionship, train employees or teach a language. An example scenario with the Furhat robot, that is also available on campus Gütersloh is shown in this video:

https://www.youtube.com/watch?v=3lEQDf9Cv4s

Despite the many possible applications, the development and use of these promising technologies presents new challenges in terms of the fit and compatibility between the needs and requirements of users and humanoid robots. One challenge lies in configuring the robots in terms of speech, facial expressions and gestures so that they meet the individual needs of the users. This is where this project comes in. The overarching goal is to configure the humanoid robot Furhat, which is already available to the Department of Engineering and Mathematics at the Gütersloh campus, so that it can interact appropriately with humans in practical application scenarios to be defined. Furhat shall first be enabled to communicate and collaborate with people in a specific practical scenario. The practical scenario will

be defined together with the project supervisors. In addition, machine learning will be used to ensure that the robot learns from interactions with its human conversation partners and remembers their language, facial expressions, gestures, preferences and conversation content. Furhat should thus "develop its skills" and be perceived as an accepted, pleasant and helpful conversation partner by its human counterparts. Alongside the technical development of Furhat, the interactions between the humans and Furhat will also be examined from a human perspective. An experimental setting will be set up to investigate specific variants of configurations with regard to Furhat's "appearance" (in particular facial expressions, language, appearance) and their effect on the human counterpart.

Brief description

At present, the importance of interaction between humans and social robots is constantly increasing. The use of social robots is becoming increasingly widespread in the healthcare sector, in education, in hotels and restaurants, in shopping malls, in industry and at home. Possible fields of application include informing and activating older people, providing support in care-related services, supporting children and young people, e.g. with autistic conditions, supervising homework and generally imparting knowledge in the education and training sector, or receiving and accompanying visitors to companies, organizations or shopping centres. Despite the many potentials of the usage of these robots, there are also numerous challenges associated with their use. For example, initial studies with humans have shown a low level of acceptance of robots among people. In order to increase acceptance, a key challenge is to configure the robots in terms of speech, facial expressions, gestures and appearance in such a way that they meet the needs of the users. This is the overarching goal of this project. A learning algorithm is to be developed that enables Furhat to remember the characteristics of the human conversation partners and adapt to them.

The main research question of the project is: How can Furhat get to know people and remember them?

To this end, the social humanoid robot head Furhat, which was developed and is marketed by the Swedish company Furhat Robotics, will be used. Furhat is a robotic head without limbs with a face projected onto the inside, where the external appearance such as skin colour, eyebrow position and make-up is adjustable. The Furhat robot thus realizes soft facial features and can participate directly in a conversation through natural movements such as shaking and nodding its head. Expressive characters are created with modern speech recognition, an advanced conversation system and automated lip synchronization.

Initially, the Furhat is available to students as a raw version. The first step is to configure it using a practical scenario (use case) and make it capable to interact with humans. In the second step, regular interactions with humans are to be carried out, e.g. weekly meetings. With the help of machine learning, Furhat should be made capable of building up a conversation history in the course of the joint meetings and thus remember past conversations with certain people. To this end, a learning algorithm is to be developed that will enable Furhat to remember and recall the language, facial expressions, gestures, preferences and content of conversations with the person in question. Furhat should therefore develop the ability, as a "conversation partner" for people, to link to information and results of past conversations and continue them. This can be done using a practical scenario (use case) in which a joint task between humans and Furhat is to be solved over several meetings. In addition, the human-robot interaction is examined from a human perspective with the help of interviews and observations.

Task

1. The raw version of the humanoid robot head Furhat has to be configured and made capable of interaction on the basis of a practical scenario (use case).
2. A learning algorithm has to be developed that enables Furhat to remember its human counterpart, his/her language, facial expressions, gestures, preferences and conversation content during regular interactions in the course of the use case developed under 1.
3. With the help of interviews with participating humans and observations of the interactions that take place, the human-robot interaction is examined from a human perspective.

Reference to the topic of data science

A machine learning algorithm will be developed.

Available resources

The Furhat robot head and the software required to program the robot are made available to the students.

Project plan

First semester: Preparation of a research exposé is an examination requirement.

Second semester: Preparation of a paper that provides an overview of the respective research area is an examination requirement.

Third semester: Preparation of a paper with initial quantitative results is an examination requirement.

Fourth semester: Master's thesis and colloquium

Suitability criteria

Mandatory: Programming skills, preferably in Python

Optional: Knowledge of conducting social science studies (e.g. conducting interviews)

Acquirable skills

- Testing a humanoid robot
- Artificial intelligence methods for problem solving
- Human-robot interaction