

# Abschlussbericht KI-Labor ITSE: KI-Labor für Methodik, Technik und Ausbildung in der IT-Systemtechnik

Jürgen Döllner, Tobias Friedrich, Bert Arnrich, Robert Hirschfeld, Christoph Lippert, Christoph Meinel





Abschlussbericht KI-Labor ITSE:  
KI-Labor für Methodik, Technik und Ausbildung in der IT-Systemtechnik



Jürgen Döllner | Tobias Friedrich | Bert Arnrich | Robert Hirschfeld |  
Christoph Lippert | Christoph Meinel

## **Abschlussbericht KI-Labor ITSE**

KI-Labor für Methodik, Technik und Ausbildung in der IT-Systemtechnik

**Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de/> abrufbar.

**Hasso-Plattner-Institut 2023**

<https://hpi.de/>

Prof.-Dr.-Helmert-Straße 2-3, 14482 Potsdam  
Tel.: +49-(0)331 5509-0 7 / Fax: +49-(0)331 5509-325  
E-Mail: [hpi-info@hpi.de](mailto:hpi-info@hpi.de)

Das Manuskript ist urheberrechtlich geschützt.

Online veröffentlicht auf dem Publikationsserver der Universität Potsdam  
URN <https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-578604>  
DOI <https://doi.org/10.25932/publishup-57860>

# Abschlussbericht KI-Labor ITSE

## Projektbezeichnung

KI-Labor für Methodik, Technik und Ausbildung in der IT-Systemtechnik zur Analyse, Planung und Konstruktion KI-basierter, komplexer IT-Systeme

## Zuwendungsempfänger:

Hasso-Plattner-Institut für Digital Engineering gGmbH ("HPI")  
Prof.-Dr.-Helmert-Str. 2-3  
14482 Potsdam

**Förderkennzeichen:** 01IS19066

**Laufzeit des Vorhabens:** 01.10.2019 – 31.12.2021

## Beteiligte Forschergruppen am HPI

- Prof. Dr. Bert Arnrich, Fachgebiet Digital Health - Connected Healthcare  
T +49 331 5509 4851  
E bert.arnrich@hpi.uni-potsdam.de
- Prof. Dr. Jürgen Döllner, Fachgebiet Computergrafische Systeme  
T +49 331 5509 171  
E juergen.doellner@hpi.uni-potsdam.de
- Prof. Dr. Tobias Friedrich, Fachgebiet Algorithm Engineering  
T +49 331 5509 411  
E tobias.friedrich@hpi.uni-potsdam.de
- Prof. Dr. Robert Hirschfeld, Fachgebiet Software-Architekturen  
T +49 331 5509 220  
E robert.hirschfeld@hpi.uni-potsdam.de
- Prof. Dr. Christoph Lippert, Fachgebiet Digital Health - Machine Learning  
T +49 331 5509 4850  
E christoph.lippert@hpi.uni-potsdam.de
- Prof. Dr. Christoph Meinel, Fachgebiet Internet-Technologien und Systeme  
T +49 331 5509 222  
E christoph.meinel@hpi.uni-potsdam.de

## Koordination

- Prof. Dr. Jürgen Döllner
- Prof. Dr. Tobias Friedrich (Stellvertreter)



# I Ausführliche Darstellung durchgeführter Arbeiten

## 1. Inhaltliche Ergebnisse

### 1.1 Arbeitsgruppe Prof. Dr. Bert Arnrich

#### 1.1.1 Arbeitspaket BA1 "Federated Learning"

Gegenstand des Arbeitspakets bildeten die Entwicklung und Evaluierung von ML-Verfahren für vertrauliche Daten, die nur eingeschränkt zentral gesammelt werden dürfen. Ein typisches Anwendungsszenario sind klinische multizentrische Studien, die häufig mit einer langwierigen und oft nur eingeschränkten zentralen Datensammlung aller beteiligten Partner einhergehen. Ziel war die Entwicklung von Methoden für verteiltes Lernen, bei denen das Lernen lokal am Ort der Datenspeicherung erfolgt, ohne dass klinische Daten exportiert und zentral gesammelt werden müssen. Ausgewählte Ergebnisse:

- Pfitzner, B., Steckhan, N., & Arnrich, B. (2021) . Federated Learning in a Medical Context: A Systematic Literature Review. ACM Transactions on Internet Technology, 21(2), 1-31. DOI: 10.1145/3412357.

Abstract: Data privacy is a very important issue. Especially in fields like medicine, it is paramount to abide by the existing privacy regulations to preserve patients' anonymity. On the other hand, data is required for research and training machine learning models that could help gain insight into complex correlations or personalised treatments that may otherwise stay undiscovered. Those models generally scale with the amount of data available, but the current situation often prohibits building large databases across sites. So it would be beneficial to be able to combine similar or related data from different sites all over the world while still preserving data privacy. Federated learning has been proposed as a solution for this, because it relies on the sharing of machine learning models (i.e. the parameters thereof), instead of the raw data itself. That means private data never leaves the site or device it was collected on. Federated learning is an emerging research area and many domains have been identified for the application of those methods. This systematic literature review provides an extensive look at the concept of and research into federated learning and its applicability for confidential healthcare datasets.

- Schmid, R., Pfitzner, B., Beilharz, J., Arnrich, B., & Polze, A. (2020). Tangle Ledger for Decentralized Learning. In Proceedings of the International Parallel and Distributed Processing Symposium Workshops, pp. 852-859. IEEE. DOI: 10.1109/IPDPSW50202.2020.00144.

Abstract: Federated learning has the potential to make machine learning applicable to highly privacy-sensitive domains and distributed datasets. In some scenarios, however, a central server for aggregating the partial learning results is not available. In fully decentralized learning, a network of peer-to-peer nodes collaborates to form a consensus on a global model without a trusted aggregating party. Often, the network consists of Internet of Things (IoT) and Edge computing nodes. Previous approaches for decentralized learning map the gradient batching and averaging algorithm from traditional federated learning to blockchain architectures. In an open network of participating nodes, the threat of adversarial nodes introducing poisoned models into the network increases compared to a federated learning scenario which is controlled by a single authority. Hence, the decentralized architecture must additionally include a machine learning-aware fault tolerance mechanism to address the increased attack surface. We propose a tangle architecture for decentralized learning, where the validity of model updates is checked as part of the basic consensus. We provide an experimental evaluation of the proposed architecture, showing that it performs well in both model convergence and model poisoning protection.

- Beilharz, J., Pfitzner, B., Schmid, R., Geppert, B., & Arnrich, A. (2021). Implicit model specialization through dag-based decentralized federated learning. In Proceedings of the 22nd International Middleware Conference, pp. 310-322. ACM. DOI: 10.1145/3464298.3493403.

Abstract: Federated learning allows a group of distributed devices to train a common machine learning model on local data. Model updates are managed either by a central entity or in a decentralized way, e.g. in a blockchain. However, this approach is not suitable for not independent and identically distributed (non-IID) data. We propose to use a directed acyclic graph (DAG) of model updates for decentralized federated learning. Instead of training a global model, in our approach clients specialize on their local data while using the model updates from other clients with similar data. Using a DAG enables the evolution of specialized models, which focus on a subset of the data and therefore cover non-IID data better than federated learning in a centralized or blockchain-based setup. To the best of our knowledge, the proposed solution is the first to unite personalization and poisoning robustness in fully decentralized federated learning. Our evaluation shows that the specialization of models emerges directly from the DAG-based communication of model updates on three different datasets. Furthermore, we show stable model accuracy and less variance across clients when compared to federated averaging.

### 1.1.2 Arbeitspaket BA2 "Internet-of-Health-Things"

Gegenstand des Arbeitspakets bildeten die Entwicklung und Evaluierung von KI-Methoden, die eine zuverlässige Erkennung von gesundheitsrelevanten Daten aus dem Alltag ermöglichen. Die Verfahren basieren auf multimodalen Sensordaten von Alltagsgeräten, die eigenständig oder über einen Sensor-Hub mit dem Internet verbunden sind (z.B. tragbare Pulsuhr, EKG Brustgurt, EEG Stirnband, WLAN-Waage, etc.). Die KI-Methoden müssen sich auf die stark schwankende Datenqualität der Alltagsdaten adaptieren können. Ausgewählte Ergebnisse:

- Zhou, L., Tunca, C., Fischer, E., Brahms, C., Ersoy, C., Granacher, U., & Arnrich, B. (2020). Validation of an IMU Gait Analysis Algorithm for Gait Monitoring in Daily Life Situations. In Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, pp. 4229-4232. IEEE. DOI: 10.1109/EMBC44109.2020.9176827.

Abstract: Gait is an essential function for humans, and gait patterns in daily life provide meaningful information about a person's cognitive and physical health conditions. Inertial measurement units (IMUs) have emerged as a promising tool for low-cost, unobtrusive gait analysis. However, large varieties of IMU gait analysis algorithms and the lack of consensus for their validation make it difficult for researchers to assess the reliability of the algorithms for specific use cases. In daily life, individuals adapt their gait patterns in response to changes in the environment, making it necessary for IMU gait analysis algorithms to provide accurate measurements despite these gait variations. In this paper, we reviewed common types of IMU gait analysis algorithms and appropriate analysis methods to evaluate the accuracy of gait parameters extracted from IMU measurements. We then evaluated stride lengths and stride times calculated from a comprehensive double integration based IMU gait analysis algorithm using an optoelectric walkway as gold standard. In total, 729 strides from five healthy subjects and three different walking patterns were analyzed. Correlation analyses and Bland-Altman plots showed that this method is accurate and robust against large variations in walking patterns (stride length: correlation coefficient ( $r$ ) was 0.99, root mean square error (RMSE) was 3% and average limits of agreement (LoA) was 6%; stride time:  $r$  was 0.95, RMSE was 4% and average LoA was 7%), making it suitable for gait evaluation in daily life situations. Due to the small sample size, our preliminary findings should be verified in future studies.

Die Bewegungsanalyse im täglichen Leben hilft bei der Überwachung des Gesundheitszustands sowohl der gesunden Bevölkerung als auch der Patienten. Für dieses Arbeitspaket wurden zwei Methoden zur Erfassung von Bewegungsdaten evaluiert: Computer Vision (CV) und Beschleunigungssensoren (IMU). Die Datenqualität, der in die

engere Wahl gekommenen kommerziellen Messgeräte, legten die Geräte für die Datenerfassung fest. Es wurden Methoden des maschinellen Lernens eingesetzt, um die Daten im Kontext der Ganganalyse, des Krafttrainings und der Erkennung menschlicher Aktivitäten auszuwerten. Ausgewählte Ergebnisse:

- Zhou, L., Tunca, C., Fischer, E., Brahms, C., Ersoy, C., Granacher, U., & Arrrich, B. (2020). Validation of an IMU Gait Analysis Algorithm for Gait Monitoring in Daily Life Situations. In Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, pp. 4229-4232. IEEE. DOI: 10.1109/EMBC44109.2020.9176827.

Abstract: Gait is an essential function for humans, and gait patterns in daily life provide meaningful information about a person's cognitive and physical health conditions. Inertial measurement units (IMUs) have emerged as a promising tool for low-cost, unobtrusive gait analysis. However, large varieties of IMU gait analysis algorithms and the lack of consensus for their validation make it difficult for researchers to assess the reliability of the algorithms for specific use cases. In daily life, individuals adapt their gait patterns in response to changes in the environment, making it necessary for IMU gait analysis algorithms to provide accurate measurements despite these gait variations. In this paper, we reviewed common types of IMU gait analysis algorithms and appropriate analysis methods to evaluate the accuracy of gait parameters extracted from IMU measurements. We then evaluated stride lengths and stride times calculated from a comprehensive double integration based IMU gait analysis algorithm using an optoelectric walkway as gold standard. In total, 729 strides from five healthy subjects and three different walking patterns were analyzed. Correlation analyses and Bland-Altman plots showed that this method is accurate and robust against large variations in walking patterns (stride length: correlation coefficient ( $r$ ) was 0.99, root mean square error (RMSE) was 3% and average limits of agreement (LoA) was 6%; stride time:  $r$  was 0.95, RMSE was 4% and average LoA was 7%), making it suitable for gait evaluation in daily life situations. Due to the small sample size, our preliminary findings should be verified in future studies.

- Zhou, L., Fischer, E., Tunca, C., Brahms, C., Ersoy, C., Granacher, U., & Arrrich, B. (2020). How We Found Our IMU: Guidelines to IMU Selection and a Comparison of Seven IMUs for Pervasive Healthcare Applications. *Sensors*, 20(15), 1-28. DOI: 10.3390/s20154090.

Abstract: Inertial measurement units (IMUs) are commonly used for localization or movement tracking in pervasive healthcare-related studies, and gait analysis is one of the most often studied topics using IMUs. The increasing variety of commercially available IMU devices offers convenience by combining the sensor modalities and simplifies the data collection procedures. However, selecting the most suitable IMU device for a certain use case is increasingly challenging. In this study, guidelines for IMU selection are proposed. In particular, seven IMUs were compared in terms of their specifications, data collection procedures, and raw data quality. Data collected from the IMUs were then analyzed by a gait analysis algorithm. The difference in accuracy of the calculated gait parameters between the IMUs could be used to retrace the issues in raw data, such as acceleration range or sensor calibration. Based on our algorithm, we were able to identify the best-suited IMUs for our needs. This study provides an overview of how to select the IMUs based on the area of study with concrete examples, and gives insights into the features of seven commercial IMUs using real data.



Abbildung: Aufstellung der Gangdatenaufzeichnung zur Bewertung des IMU-Ganganalysealgorithmus. Links: der 10 m lange OptoGait-Gang. Das OptoGait-System (Microgait, Italien) wurde als Goldstandard zur Kreuzvalidierung der aus IMU-Daten berechneten Gangparameter verwendet. Rechts: Die IMUs wurden oben auf den Schuhen befestigt, um Gangdaten aufzuzeichnen.

- Albert, J., Owolabi, V., Gebel, A., Brahms, C., Granacher, U., & Arnrich, B. (2020). Evaluation of the Pose Tracking Performance of the Azure Kinect and Kinect v2 for Gait Analysis in Comparison with a Gold Standard: A Pilot Study. *Sensors*, 20(18), DOI: 10.3390/s20185104.

Abstract: Gait analysis is an important tool for the early detection of neurological diseases and for the assessment of risk of falling in elderly people. The availability of low-cost camera hardware on the market today and recent advances in Machine Learning enable a wide range of clinical and health-related applications, such as patient monitoring or exercise recognition at home. In this study, we evaluated the motion tracking performance of the latest generation of the Microsoft Kinect camera, Azure Kinect, compared to its predecessor Kinect v2 in terms of treadmill walking using gold standard Vicon multi-camera motion capturing system and the 39 marker Plug-in Gait model. Five young and healthy subjects walked on a treadmill at three different velocities while data were recorded simultaneously with all three camera systems. An easy-to-administer camera calibration method developed here was used to spatially align the 3D skeleton data from both Kinect cameras and the Vicon system. With this calibration, the spatial agreement of joint positions between the two Kinect cameras and the reference system was evaluated. In addition, we compared the accuracy of certain spatio-temporal gait parameters, i.e., step length, step time, step width, and stride time calculated from the Kinect data, with the gold standard system. Our results showed that the improved hardware and the motion tracking algorithm of the Azure Kinect camera led to a significantly higher accuracy of the spatial gait parameters than the predecessor Kinect v2, while no significant differences were found between the temporal parameters. Furthermore, we explain in detail how this experimental setup could be used to continuously monitor the progress during gait rehabilitation in older people.

- Rahn, V. X., Zhou, L., Klieme, E., & Arnrich, B. (2021). Optimal sensor placement for human activity recognition with a minimal smartphone–IMU setup. In *SENSORNETS*, pp. 37-48.

Abstract: Human Activity Recognition (HAR) of everyday activities using smartphones has been intensively researched over the past years. Despite the high detection performance, smartphones can not continuously provide reliable information about the currently conducted activity as their placement at the subject's body is uncertain. In this study, a system is developed that enables real-time collection of data from various Bluetooth inertial measurement units (IMUs) in addition to the smartphone. The contribution of this work is an extensive overview of related work in this field and the identification of unobtrusive, minimal combinations of IMUs with the smartphone that achieve high recognition performance. Eighteen young subjects with unrestricted mobility were recorded conducting seven daily-life activities with a smartphone in the pocket and five IMUs at different body positions. With a

Convolutional Neural Network (CNN) for activity recognition, activity classification accuracy increased by up to 23% with one IMU additional to the smartphone. An overall prediction rate of 97% was reached with a smartphone in the pocket and an IMU at the ankle. This study demonstrated the potential that an additional IMU can improve the accuracy of smartphone-based HAR on daily-life activities.

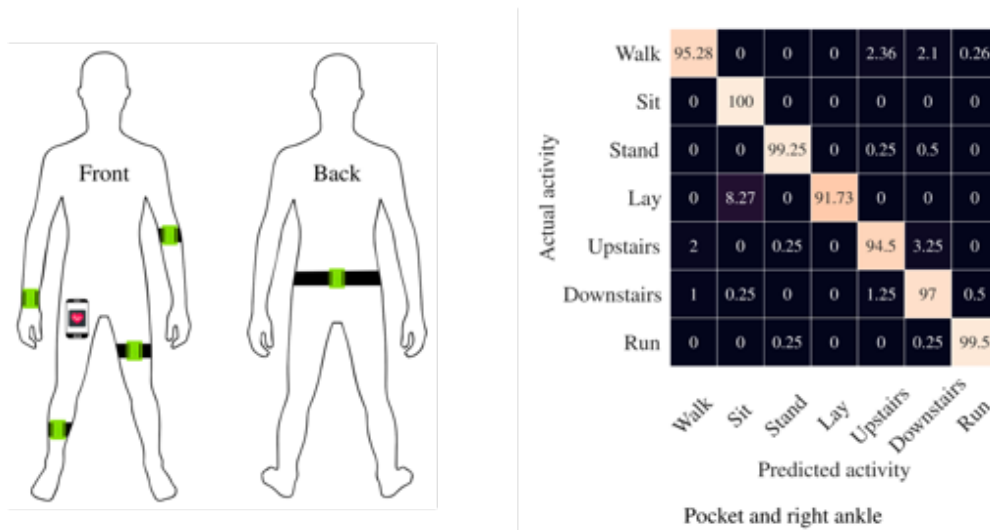


Abbildung: Links: Sensoraufbau. Die fünf IMUs wurden am linken Oberschenkel, rechten Knöchel, linken Arm, rechten Handgelenk und unteren Rücken platziert. Das Smartphone mit der Aufzeichnungs-App wurde in der rechten Tasche platziert. Rechts: Konfusionsmatrix mit Klassifizierungsergebnissen vom rechten Knöchel (einer der optimalen Sensorpositionen).

### 1.1.3 Arbeitspaket BA-3 "Connected Health Patientenakte"

Gegenstand des Arbeitspakets war die Fusion von elektronischer Patientenakte mit gesundheitsrelevanten Alltagsdaten. Dadurch wird die Gestaltung von KI-Modellen ermöglicht, die insbesondere zur Prävention chronischer Erkrankungen, zur Evaluierung von Therapien und zur frühzeitigen Erkennung von Gesundheitsrisiken eingesetzt werden können. Es wurden KI-Modelle entwickelt, die sowohl Daten aus vorhandenen elektronischen Patientenakten als auch Sensordaten aus dem Alltag einbeziehen können. Ausgewählte Ergebnisse:

- Pfitzner, B., Steckhan, N., & Arrich, B. (2022). Federated Learning in a Medical Context: A Systematic Literature Review. To appear in Transactions on Internet Technology.

Abstract: Data privacy is a very important issue. Especially in fields like medicine, it is paramount to abide by the existing privacy regulations to preserve patients' anonymity. On the other hand, data is required for research and training machine learning models that could help gain insight into complex correlations or personalised treatments that may otherwise stay undiscovered. Those models generally scale with the amount of data available, but the current situation often prohibits building large databases across sites. So it would be beneficial to be able to combine similar or related data from different sites all over the world while still preserving data privacy. Federated learning has been proposed as a solution for this, because it relies on the sharing of machine learning models, instead of the raw data itself. That means private data never leaves the site or device it was collected on. Federated learning is an emerging research area and many domains have been identified for the application of those methods. This systematic literature review provides an extensive look at the concept of and research into federated learning and its applicability for confidential healthcare datasets.

- Schmid, R., Pfitzner, B., Beilharz, J., Arrrich, B. & Polze, A. (2020). In Proceedings of the International Parallel and Distributed Processing Symposium Workshops, pp. 852-859. IEEE. DOI: 10.1109/IPDPSW50202.2020.00144.

Abstract: Federated learning has the potential to make machine learning applicable to highly privacy-sensitive domains and distributed datasets. In some scenarios, however, a central server for aggregating the partial learning results is not available. In fully decentralized learning, a network of peer-to-peer nodes collaborates to form a consensus on a global model without a trusted aggregating party. Often, the network consists of Internet of Things (IoT) and Edge computing nodes. Previous approaches for decentralized learning map the gradient batching and averaging algorithm from traditional federated learning to blockchain architectures. In an open network of participating nodes, the threat of adversarial nodes introducing poisoned models into the network increases compared to a federated learning scenario which is controlled by a single authority. Hence, the decentralized architecture must additionally include a machine learning-aware fault tolerance mechanism to address the increased attack surface. We propose a tangle architecture for decentralized learning, where the validity of model updates is checked as part of the basic consensus. We provide an experimental evaluation of the proposed architecture, showing that it performs well in both model convergence and model poisoning protection.

- Steckhan, N., & Arrrich, B. (2020). Quantified Complementary and Alternative Medicine: Convergence of Digital Health Technologies and Complementary and Alternative Medicine. *Complement Med. Res.*, 27(3), 131-133. DOI: 10.1159/000506672.

Abstract: In recent years, consumer health technologies including wearables, Internet-of-Things, and health apps that support wellness and healthy lifestyle have emerged rapidly. However, only a few of these new technologies have been certificated as medical product yet (e.g., only 40 CE-certified apps in Germany as of July 2019). Common application scenarios include lifestyle-associated aspects like diet, exercise, stress management, or combinations of it. Interestingly, the same aspects are addressed by several fields of complementary and alternative medicine (CAM) like integrative or Mind Body Medicine. This synergy has mainly two reasons, first the modern technologies are getting better to process lifestyle data, and second, the regulations (Medical Device Regulation) are more complex if one includes non-wellness health claims (e.g., dosing of medications, medical decisions). In order to apply such technologies in medicine, however, methods are needed that are also suitable for research beyond big players and big data (e.g., Apple, Google, and Samsung), such as in complementary medical research with small numbers of cases. The support from validated models and tools can be a beneficial ingredient if not transforming complementary medicine research (think of the one channel ECG in Apple Watch). For example, monitoring vital signs in real time and in daily life can dramatically enhance the field of research.

- Morassi Sasso, A., Datta, S., Jeitler, M., Steckhan, N., Kessler, C., Michalsen, A., Arrrich, B., & Böttinger, E. (2020). HYPE: Predicting Blood Pressure from Photoplethysmograms in a Hypertensive Population. In Proceedings of the International Conference on Artificial Intelligence in Medicine, pp. 325-335. Springer. DOI: 10.1007/978-3-030-59137-3\_29.

Abstract: The state of the art for monitoring hypertension relies on measuring blood pressure (BP) using uncomfortable cuff-based devices. Hence, for increased adherence in monitoring, a better way of measuring BP is needed. That could be achieved through comfortable wearables that contain photoplethysmography (PPG) sensors. There have been several studies showing the possibility of statistically estimating systolic and diastolic BP (SBP/DBP) from PPG signals. However, they are either based on measurements of healthy subjects or on patients on (ICUs). Thus, there is a lack of studies with patients out of the normal range of BP and with daily life monitoring out of the ICUs. To address this, we created a dataset (HYPE) composed of data from hypertensive subjects that executed a stress test and had 24-h monitoring. We then trained and compared machine learning (ML) models to predict BP. We evaluated handcrafted feature extraction approaches vs image representation ones and compared different ML algorithms for both. Moreover, in order to evaluate the models in a different scenario, we used an openly available set from a stress test with healthy subjects (EVAL). The best results for our HYPE dataset were in the stress test and had a mean absolute error (MAE) in mmHg of 8.79 ( $\pm 3.17$ ) ( $\pm 3.17$ ) for SBP and 6.37 ( $\pm 2.62$ ) ( $\pm 2.62$ ) for DBP; for our EVAL dataset it was 14.74 ( $\pm 4.06$ ) ( $\pm 4.06$ ) and 7.12 ( $\pm 2.32$ ) ( $\pm 2.32$ ) respectively. Although having tested

a range of signal processing and ML techniques, we were not able to reproduce the small error ranges claimed in the literature. The mixed results suggest a need for more comparative studies with subjects out of the intensive care and across all ranges of blood pressure. Until then, the clinical relevance of PPG-based predictions in daily life should remain an open question.

- Moontaha, S., Steckhan, N., Kappattanavar, A., Surges, R., & Arrrich, B. (2020). Self-prediction of seizures in drug resistance epilepsy using digital phenotyping: a concept study. In Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare. pp. 384-387. EAI. DOI: 10.1145/3421937.3421943.

Abstract: Drug-resistance is a prevalent condition in children and adult patients with epilepsy. The quality of life of these patients is profoundly affected by the unpredictability of seizure occurrence. Some of these patients are capable of reporting self-prediction of their seizures by observing their affectivity. Some patients report no signs of feeling premonitory symptoms, prodromes, or aura. In this paper, we propose a concept study that will provide objective information to self-predict seizures for both the patient groups. We will develop a model using digital phenotyping which takes both ecological momentary assessment and data from sensor technology into consideration. This method will be able to provide a feedback of their premonitory symptoms so that a pre-emptive therapy can be associated to reduce seizure frequency or eliminate seizure occurrence.

### Stationäres Monitoring:

*Mit Sensordaten angereicherte klinische Phänotypisierung für die integrative Medizin*

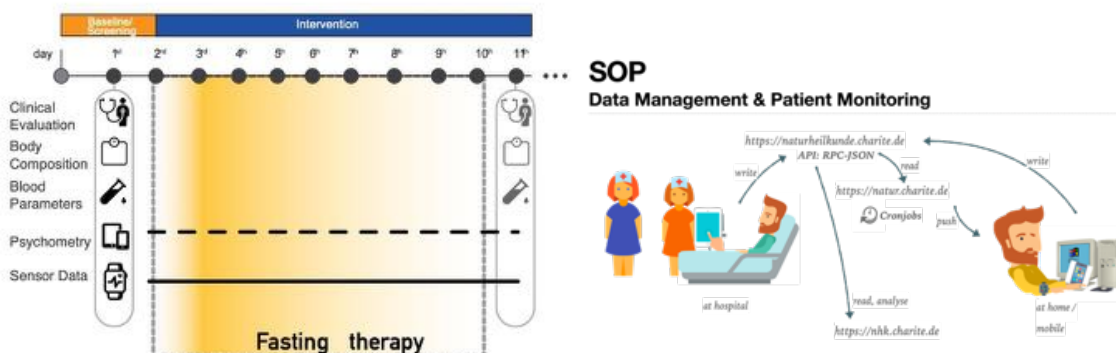
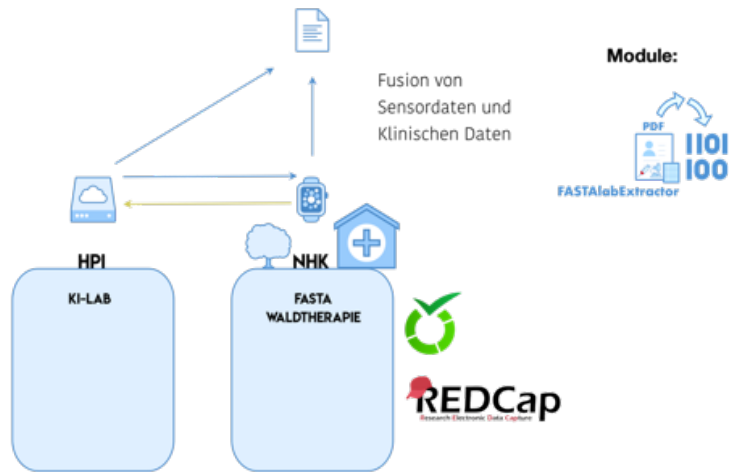


Abbildung: Studiendesign (links). Vor der Baseline wurden die Teilnehmer mit einem Online Survey gescreent. Beim ersten Besuch (bzw. Eine Sensorbox nach Hause versandt) wurden Fragebögen, Vitaldaten und Blut abgenommen. Den Teilnehmern wurden außerdem Sensoren angelegt, sie erhielten eine Actigraph (bzw. Empatica oder Everion Armband) und wurden in der Bedienung der Offline Survey App geschult. Die Teilnehmer trugen die Sensoren und protokollierten ihre subjektiven Zustände mittels validierter Fragebögen. Nach Abschluss der Interventionszeit wurden die Teilnehmer angewiesen, weitere Fragebögen bis nach 3, 6 und 12 Monaten auszufüllen. Datenfluss (rechts). Mittels 3 dedizierter VM werden klinische Daten sowohl auf Station als auch ambulant gesammelt, gemonitort und analysiert.



Oben dargestellt ist die Verbindung zwischen Ressourcen des klinischen Partners (Charité) und dem HPI. Die klinischen Daten können so auf Seiten der Charité bleiben und ausschließlich Pseudonyme (token) werden für spätere Integration mit Sensordaten auf HPI Server verschickt und prozessiert wieder aggregiert zurückgesandt.

Modul FASTA labExtractor: Um klinische Labordaten von stationären Patienten zu erfassen, sind wir nun in der Lage, die Daten nach der Erstellung des Labor-PDFs einfach zu verarbeiten. Obwohl dieser halbautomatische Prozess durch direkte Exporte oder Anbindung an die Routinelabordatenbank (LADR) verbessert werden könnte, sind wir nicht mehr auf die manuelle Übertragung in unsere Forschungsdatenbank angewiesen. Nachfolgend ein Auszug der Routinelabor Daten von 440 Krankenhauspatienten.

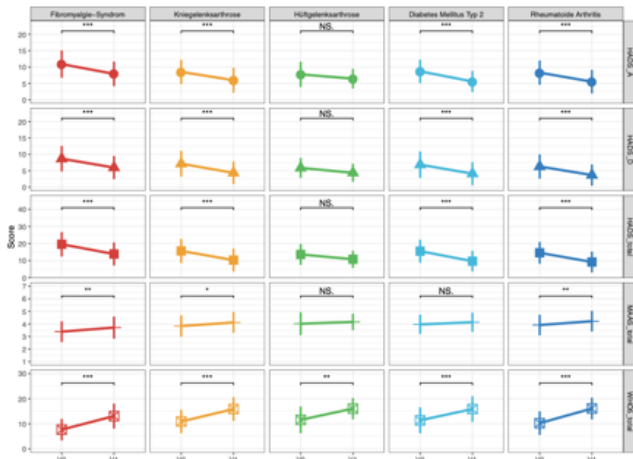


Abbildung: Links: Auswirkungen des stationären Aufenthaltes (integrative Fastenintervention) auf psychologische Dimensionen. (HADS A: Ängste, HADS D: Depressionen, MAAS: Achtsamkeit, WHO5: Wohlbefinden) Rechts: Auswirkungen des stationären Fastens auf Blutparameter.

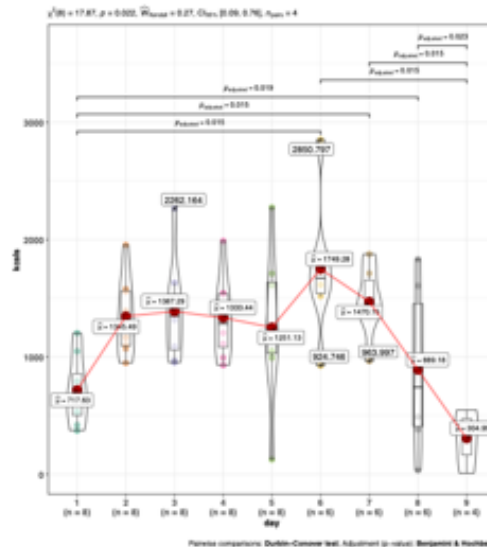


Abbildung: Actigraphie Resultate zeigen signifikante Verbesserung des Energieumsatzes nach Tag 6. Dies reflektiert die positiven psychometrischen Veränderungen nach dem Fasten.

**Ambulantes Monitoring:**  
 Vorstationäres Monitoring und Sensordatenerfassung

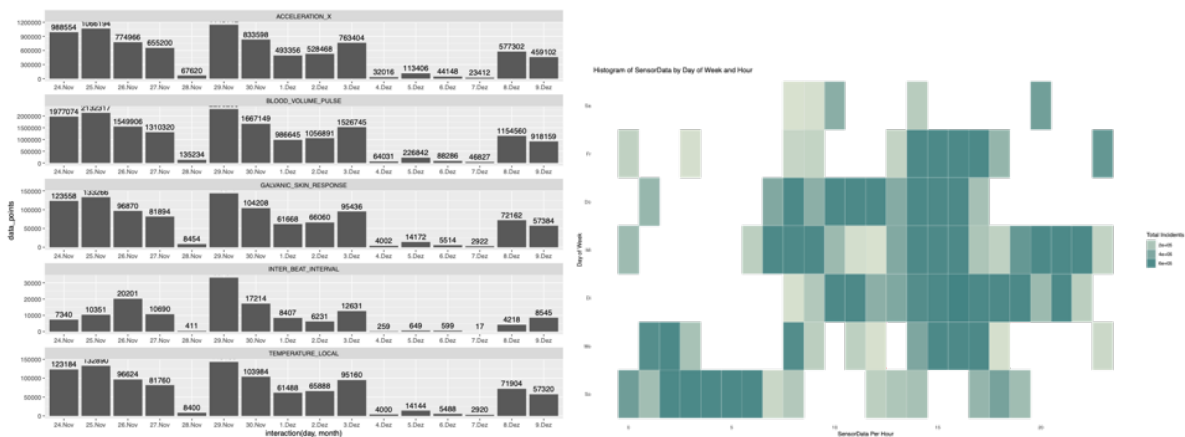


Abbildung: Vorstationäres Monitoring für die Generierung von Baseline Daten aus Sensoren: FASTAvital Kohorte (links): Ausbeute der Daten einer vorstationären Patient\*in (fastavital001) via SensorHub (Empatica E4) Rechts: Aggregation der Sensordaten für 24/7 monitoring.

**Ambulantes Monitoring: Naturotherapie**  
 Entwurf einer Pilotstudie für neurourbanistisches Monitoring - Sensoren im Feld

Zusammen mit der Charité Naturheilkunde wurden die Effekte von Naturexposition bei verschiedenen Erkrankungen erforscht. Neben möglichen Lebensstil-Faktoren wie mangelnder Bewegung sind Erkrankungen wie metabolisches Syndrom oder psychoonkologische Begleiterkrankungen auch mit Stress assoziiert. Im Projekt wurden dabei Methoden des Mobile Sensing entwickelt und verwendet, um Bedingungen, welche sich aus Standort und Umweltdaten ableiten lassen, mit Vitalparametern zu integrieren und zu analysieren. Anhand dieses Beispiels der Naturtherapie versuchten wir, systematische

Erkenntnisse aus Pilotversuchen zu gewinnen. Pilotversuche sind essentiell für die Bewertung und Entscheidungsfindung von methodischen Ansätzen in Bezug auf Kombinationen von verschiedenen Studienobjekten (z.B. Menschen, Sensorgeräte, Studieninterventionen). Wir präsentierten ein systematisches Bewertungsverfahren zur Optimierung der Ausbeute und Struktur einer Studie unter Verwendung des Konzepts des Designraums. Optimierungsverfahren können durchgeführt werden, um die Diskriminationsgenauigkeit verschiedener Studienbedingungen auf individueller oder durchschnittlicher Ebene anhand von Serien von Einzelfallversuchen zu erhöhen. Der Stand der Technik des Studiendesigns in der evidenzbasierten Medizin sind randomisierte kontrollierte Studien. Die Anwendung solcher Designs (z.B. statistische Modelle mit longitudinalen wiederholten Messungen) in einem mobilen Gesundheitskontext kann nicht wie erwartet funktionieren, wenn man mit kontinuierlichen Daten von Sensoren arbeitet. Daher hilft unser Ansatz den Forschern, groß angelegte, longitudinale Fernstudien durchzuführen, die nicht nur selbstberichtete Ergebnisse, sondern tatsächliche Messungen von funktionalen und verhaltensbezogenen Ergebnissen erfassen.

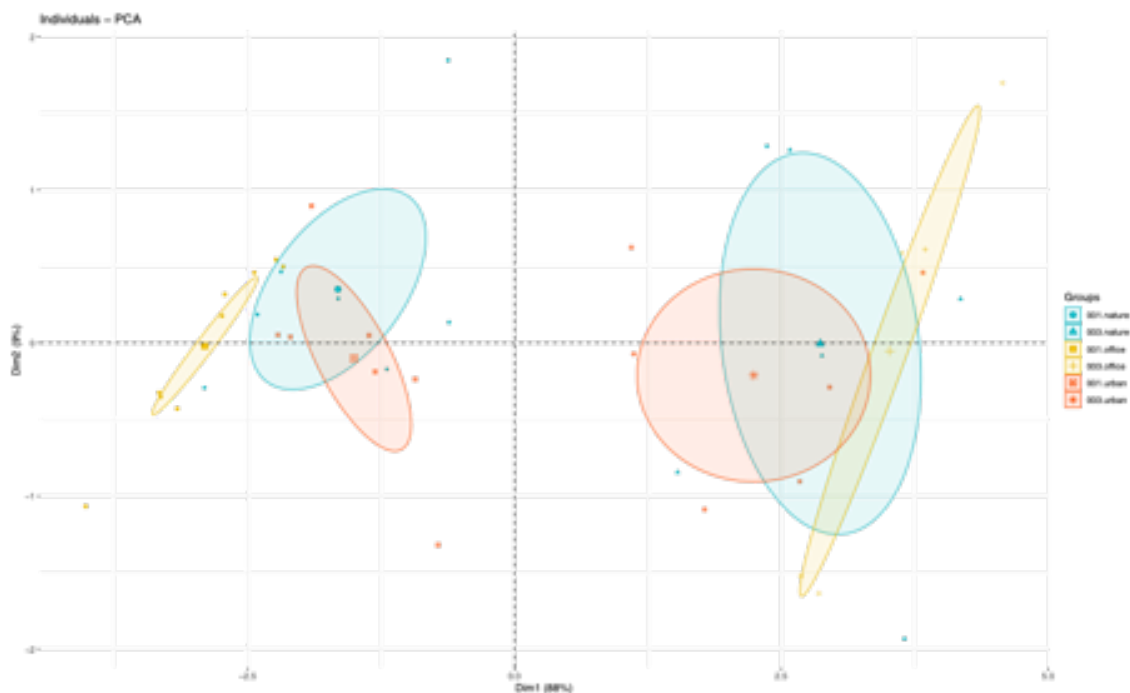


Abbildung: PCA-Plot von Personen unter 3 Bedingungen (Büro, Natur, Stadt), um die Verteilung der Variabilität, den Beitrag von HRV- time domain features und die Unterscheidungsleistung bei Verwendung des Corsano-Armbandes zu bewerten.

#### 1.1.4 Arbeitspaket BA-4 "Transparente KI"

Gegenstand des Arbeitspakets bildete die Entwicklung von Verfahren zur verständlichen Darstellung der Arbeitsweise von maschinellen Lernverfahren. Damit sollen insbesondere Ärzte und Patienten in die Lage versetzt werden, maschinelle Lernverfahren besser zu verstehen. Dabei soll der komplette Datenpfad von der Vorverarbeitung über die Modellierung hin zur Inferenz einbezogen werden.

#### Körperliches Training mit Hilfe von visualisierten Bewegungen

Die Wirkung von körperlicher Bewegung, insbesondere im Rahmen der Rehabilitation, hängt von der Präzision der Bewegungen ab. Es wurde ein Heimtrainings-System entwickelt (entstanden aus einem Master-Studentenprojekt), bei dem die Bewegungsinformationen mithilfe eines Convolutional Neural Network (CNN) extrahiert und die Bewegungen auf einem interaktiven Dashboard visualisiert werden.

Die bisherigen Ergebnisse wurden in Form eines wissenschaftlichen Fachbeitrags aufbereitet:

- Albert, J., Zhou, L., Glöckner, P., Trautmann, J., Ihde, L., Eilers, J., Kamal, M., & Arnrich, B. (2020). Will You Be My Quarantine: A Computer Vision and Inertial Sensor Based Home Exercise System. In Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare, pp. 380-383. EAI. DOI: 10.1145/3421937.3421946.

Abstract: The quarantine situation inflicted by the COVID-19 pandemic has left many people around the world isolated at home. Despite the large variety of mobile device-based self exercise tools for training plans, activity recognition or repetition counts, it remains challenging for an inexperienced person to perform fitness workouts or learn a new sport with the correct movements at home. As a proof of concept, a home exercise system has been developed in this contribution. The system takes computer vision and inertial sensor data recorded for the same type of exercise as two independent inputs, and processes the data from both sources into the same representations on the levels of raw inertial measurement unit (IMU) data and 3D movement trajectories. Moreover, a Key Performance Indicator (KPI) dashboard was developed for data import and visualization. The usability of the system was investigated with an example use case where the learner equipped with IMUs performed a kick movement and was able to compare it to that from a coach in the video.

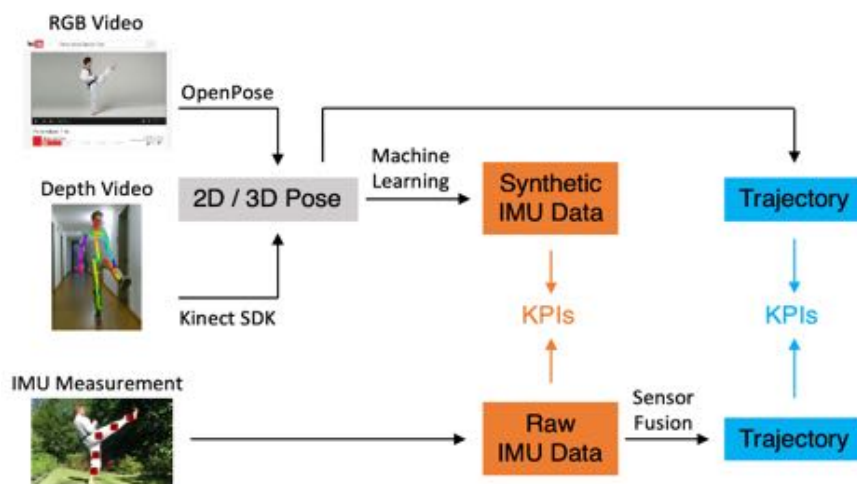


Abbildung: Systemarchitektur. Das System extrahiert 2D/3D-Positionsdaten aus einem gegebenen Eingangsvideo, um Informationen über die Trajektorie der Gelenke zu erhalten. Außerdem wird ein neuronales Netzwerk verwendet, um synthetische IMU-Daten aus 2D-Positionen vorherzusagen, die aus dem Eingangsvideo gewonnen werden. Wenn der Benutzer die Übungen zu Hause mit angeschlossenen IMUs wiederholt, wird die 3D-Trajektorie aus den IMU-Daten berechnet. Durch die Definition von KPIs für jede der beiden Darstellungen können detaillierte Einblicke in die Leistung der ausgeführten Übung gegeben werden.

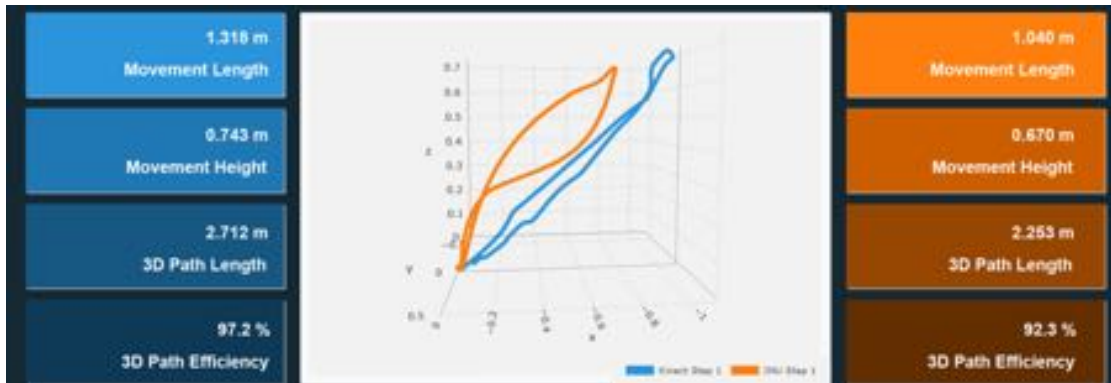


Abbildung: Screenshot des Trajektorienabschnitts auf dem KPI-Dashboard mit Daten von einem Fuß. 3D-Trajektorien und abgeleitete KPIs vom Trainer mit Computer-Vision-Daten (blau) und dem Lerner mit IMU-Daten (orange) konnten deutlich visualisiert werden.

### Kontinuierliches Glukose Monitoring und circadiane Biomarker

Integrative Analysen von kombinierten Daten aus mobilen Apps, Wearables, Fragebögen, metabolischen und genetischen Analysen unter Verwendung von temporalem Mining und geeignetem maschinellem Lernen werden durchgeführt, um die optimale Tool/Marker-Kombination für die Entwicklung eines personalisierten Chrononährstoffplans aufzudecken. Eine Validierung der Bedeutung hybrider Biomarker (statisch und dynamisch) unter Verwendung von chronotyp-bezogenem Merkmalsextraktion und prädiktiver Modellierung ist gewünscht. Wir entwickelten die Möglichkeit einer kontinuierlichen Chronotyp-Vorhersage unter Verwendung einer verschmolzenen Verarbeitung von dynamischen zirkadianen Merkmalen aus Sensordaten (Glukosesensoren, Aktigraphen) und subjektiven Momentaufnahmen. Hierfür ist es auch entscheidend, in einem interaktiven Austausch mit den medizinischen Domain-Expert\*innen oder später Patient\*innen zu verstehen, welche Merkmale für die Modelle prädiktiv sind.

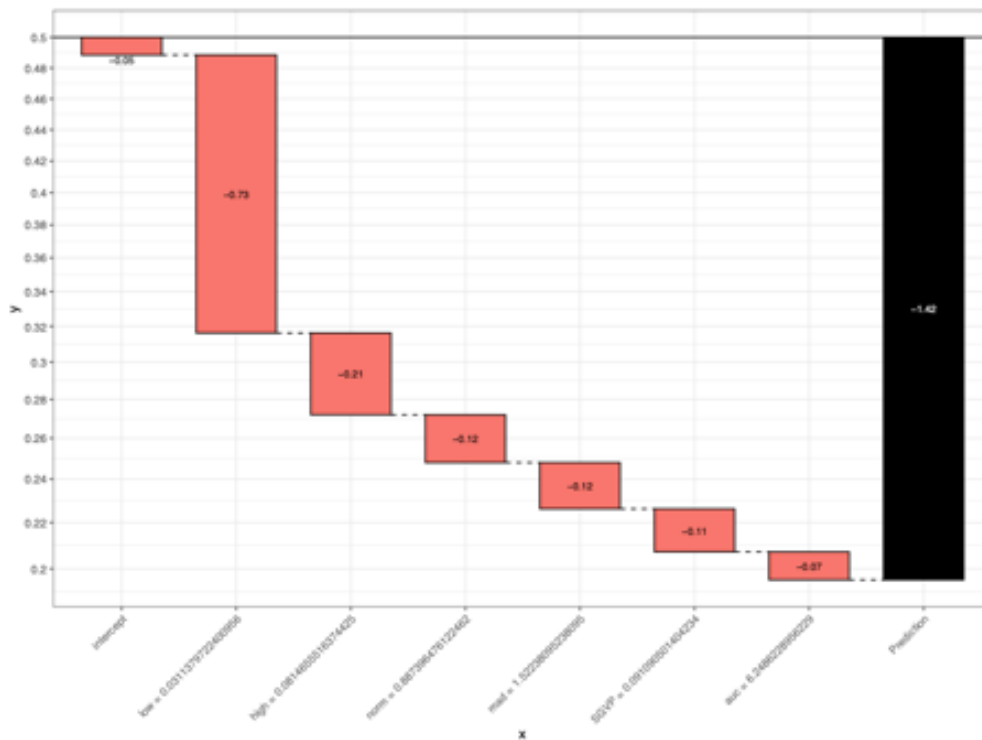
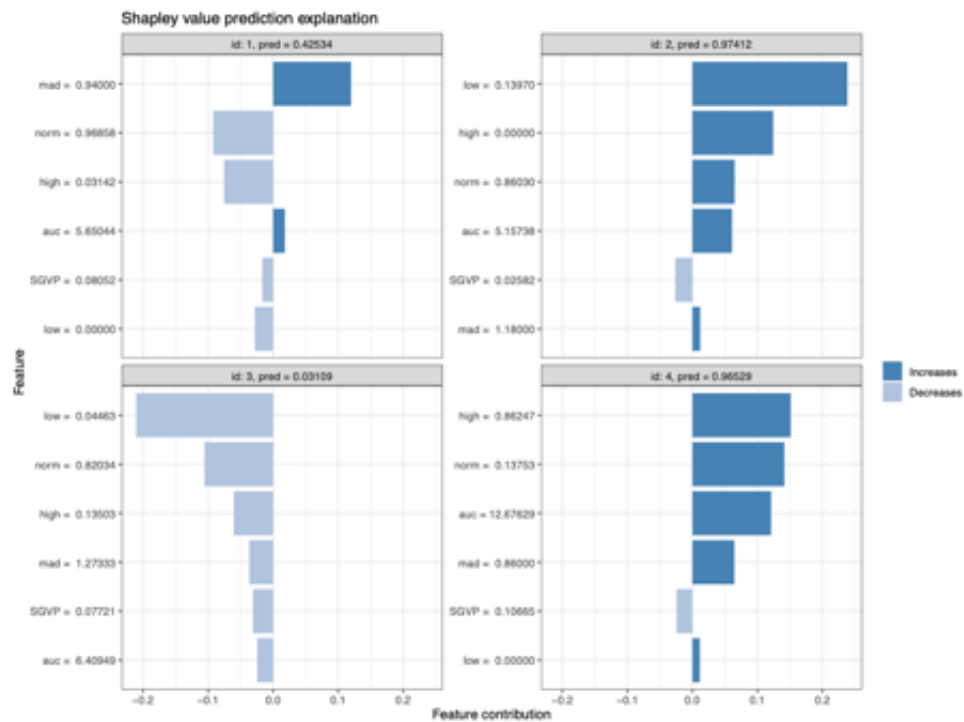


Abbildung: Model-spezifische Erklärung am Beispiel einer binären Klassifikation von kontinuierlichen Glucose Monitoring (CGMs) Systemen durch die XGBoost Methode. Die Abbildung zeigt die Erklärung der log-odds-Modellvorhersage von -1.42 (die y-Achse zeigt die Wahrscheinlichkeit, die Balkenbeschriftungen zeigen den log-odds-Einfluss jeder Variable)

In Diskussion mit den Patient\*innen (Diabetes Typ 1) konnte gezeigt werden, dass das Modell die richtigen globalen Merkmale für die Klassifizierung verwendet. Weiter wurde evaluiert, ob und wie spezifische und individuelle Unterschiede mittels Modell agnostischer Methoden dargestellt sind.



Somit können Forscher\*innen, medizinisches Personal oder die Patient\*innen selbst erkennen, welche individuellen z.B. anhand der Tageszeit differenzierten Merkmale für die Klassifikationsmethode entscheidend waren. Damit erhoffen wir, mehr Vertrauen in solche Methoden gewinnen zu können. Es ist hierbei dennoch unerlässlich, die Aus- und Weiterbildung von Mediziner\*innen und anderen Gesundheitsberufen in interdisziplinären Diskursen und deren kritischen Umgang zu stärken.

Für die Implementierung, Erprobung und Auswertung der Arbeitspakete BA-1, BA-2, BA-3 und BA-4 wurden die spezialisierte Hardware und Speicherkapazität (ML-/DL-Hardware, GPUs und FPGAs) eingesetzt, die über die KI-Labor-Rechentechnik bereitgestellt war. Plangemäss wurde darüber hinaus arbeitspaketspezifische apparative Ausstattung (insbesondere Kleingeräte) beschafft.

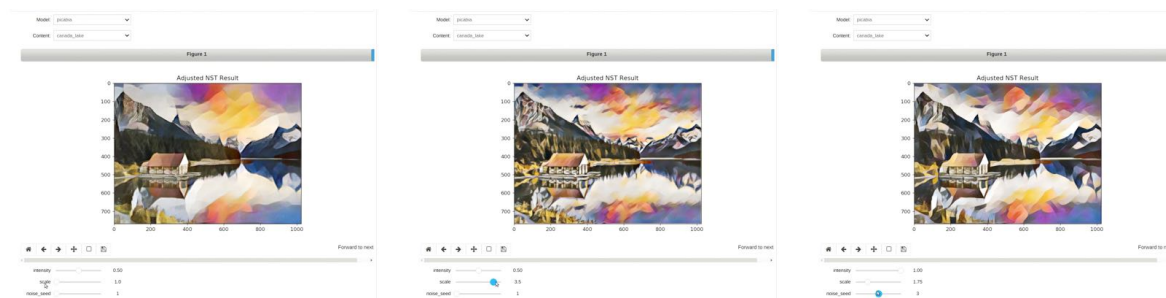
## 1.2 Arbeitsgruppe Prof. Jürgen Döllner

### 1.2.1 Arbeitspaket JD1 "KI-basierte Neural Style Transfer Verfahren"

Gegenstand des Arbeitspakets bildeten KI-Verfahren der Kategorie "Neural Style Transfer". Ihr Ziel ist es, zu einem gegebenen digitalen Eingabebild ein graphisch transformiertes Bild abzuleiten. Das Wirkprinzip ist dabei, Struktur- und Feature-Elemente des Eingabebilds zu berechnen und diese durch Elemente zu ersetzen, die aus einem Trainingsbilddatensatz stammen. Die Transformation beruht ausschließlich auf der Substitution grafischer Elemente. Anschaulich lassen sich damit grafische Stile, die z.B. durch klassische Bildern repräsentiert sind, für individuelle Bilder imitieren. Die bisherigen Ergebnisse umfassen:

Sogenannte Feed-Forward-Convolutional-Neural-Networks erfreuen sich bei Bildstilisierung einer großen Beliebtheit. Ist der Stil allerdings einmal trainiert, kann dieser nicht verändert bzw. angepasst werden. Arbitrary-Style-Transfer-Netzwerke ermöglichen die Übertragung beliebiger Bilder über ein einziges Netzwerk, allerdings sind hier die Resultate nicht auf einer fein-granularen Ebene kontrollierbar und die Übertragung kann komplexe Stilmerkmale nicht reproduzieren. Darüber hinaus wird der Transfer in praktischen Anwendungen in der Regel auf hochauflösende Bilder angewendet. Jedoch ist während des Trainings die trainierbare Inhalts- und Stilgröße aufgrund des hohen Speicherverbrauchs des jeweiligen Merkmalsextraktionsnetzwerks stark eingeschränkt. Dies führt dazu, dass in hochauflösenden Ausgabebildern sehr kleinskalige Stilelemente erzeugt werden und große Stilelemente nicht reproduzierbar sind.

Um die obigen Limitierung auszugleichen wurden sogenannte Two-Stream-Netzwerke erforscht, die eine nahtlose Kontrolle der Stilskalen ermöglichen und die Laufzeit und Speichernutzung verbessern sowie in der Lage sind, große Stilelemente auf hochauflösenden Bildern zu erzeugen. Auf der ursprünglichen Bildauflösung wird nur mit einigen wenigen "flachen" Netzwerk-Schichten gearbeitet und die Mehrheit der Schichten (Transformatorschichten) werden auf einer geringeren Auflösung des Bildes angewendet. Die resultierenden Ströme werden durch ein bilaterales Upsampling-Modul im Merkmalsraum kombiniert, wobei die Merkmale mit niedriger Auflösung auf die ursprüngliche Bildmerkmalsauflösung extrapoliert werden. Es wurde ein softwaretechnischer Framework und verschiedene Demonstratoren entwickelt, das es ermöglicht verschiedene Variationen der Kombination der Ströme im Merkmalsraum zu evaluieren sowie Lastentests durchzuführen.



Beispiele eines Demonstrators für das KI-basierte Two-Stream Neural Style Transfer Verfahren.

Für die praktischen Arbeiten, insbesondere zum Test, Integration und Evaluierung des Frameworks und für die Demonstratoren, wurde hochspezialisierte Hardware und Speicherkapazität des KI-Labors herangezogen. Sehr erfolgreich konnten Trainings- und Testdaten dazu verwendet werden, um die neu entwickelten Verfahren für praxisnahe, realistische Szenarien testen zu können.

Die Bearbeitung des Arbeitspaket wurde plangemäss durchgeführt; die wissenschaftlichen Ergebnisse wurden in den folgenden Publikationen zusammengefasst:

- Shekhar, S., Reimann, M., Mayer, M., Semmo, A., Pasewaldt, S., Döllner, J., & Trapp, M. (2021). Interactive Photo Editing on Smartphones via Intrinsic Decomposition. *Computer Graphics Forum*, 40(2), 497-510, DOI: 10.1111/cgf.142650.

Abstract: Intrinsic decomposition refers to the problem of estimating scene characteristics, such as albedo and shading, when one view or multiple views of a scene are provided. The inverse problem setting, where multiple unknowns are solved given a single known pixel-value, is highly under-constrained. When provided with correlating image and depth data, intrinsic scene decomposition can be facilitated using depth-based priors, which nowadays is easy to acquire with high-end smartphones by utilizing their depth sensors. In this work, we present a system for intrinsic decomposition of RGB-D images on smartphones and the algorithmic as well as design choices therein. Unlike state-of-the-art methods that assume only diffuse reflectance, we consider both diffuse and specular pixels. For this purpose, we present a novel specular extraction algorithm based on a multi-scale intensity decomposition and chroma inpainting. At this, the diffuse component is further decomposed into albedo and shading components. We use an inertial proximal algorithm for non-convex optimization (iPiano) to ensure albedo sparsity. Our GPU-based visual processing is implemented on iOS via the Metal API and enables interactive performance on an iPhone 11 Pro. Further, a qualitative evaluation shows that we are able to obtain high-quality outputs. Furthermore, our proposed approach for specular removal outperforms state-of-the-art approaches for real-world images, while our albedo and shading layer decomposition is faster than the prior work at a comparable output quality. Manifold applications such as recoloring, retexturing, relighting, appearance editing, and stylization are shown, each using the intrinsic layers obtained with our method and/or the corresponding depth data.

- Masopust, L., Pasewaldt, S., Döllner, J., & Trapp, M. (2021). Integration of Image Processing Techniques into the Unity Game Engine. In *Proceedings of the 1st International Conference on Image Processing and Vision Engineering*, pp. 137-144.

Abstract: This paper describes an approach for using the Unity game engine for image processing by integrating a custom GPU-based image processor. For, it describes different application levels and integration approaches for extending the Unity game engine. It further documents the respective software components and implementation details required, and demonstrates use cases such as scene post-processing and material-map processing.

- Fregien, F., Pasewaldt, S., Döllner, J., & Trapp, M. (2021). Service-based Processing of Gigapixel Images. In *Proceedings of the 29th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*.

Abstract: With the improvement of cameras and smartphones, more and more people can now take high-resolution pictures. Especially in the field of advertising and marketing, images with extremely high resolution are needed, e. g., for high quality print results. Also, in other fields such as art or medicine, images with several billion pixels are created. Due to their size, such gigapixel images cannot be processed or displayed similar to conventional images. Processing methods for such images have high performance requirements. Especially for mobile devices, which are even more limited in screen size and memory than computers, processing such images is hardly possible. In this thesis, a service-based approach for processing gigapixel images is presented to approach this problem. Cloud-based processing using different microservices enables a hardware-independent way to process gigapixel images. Therefore, the concept and implementation of such an integration into an existing service-based architecture is presented. To enable the exploration of gigapixel images, the integration of a gigapixel image viewer into a web application is presented. Furthermore, the design and implementation will be evaluated with regard to advantages, limitations, and runtime.

- Buchheim, B., Reimann, M., Pasewaldt, S., Döllner, J., & Trapp, M. (2021). StyleTune: Interactive Style Transfer Enhancement on Mobile Devices. In *Proceedings of the 21st ACM SIGGRAPH Appy Hour*, ACM. DOI: 10.1145/3450415.3464400.

Abstract: We present StyleTune, a mobile app for interactive multi-level control of neural style transfers that facilitates creative adjustments of style elements and enables high output fidelity. In contrast to current mobile neural style transfer apps, StyleTune supports users to adjust both the size and orientation of style elements, such as brush strokes and texture patches, on a global as well as local level. To this end, we use stroke-adaptive neural networks to

train and seamlessly integrate different stroke sizes at run-time. For it, two neural-network architectures for stroke-adaptive style transfer are implemented: the first excels at fine-grained stroke-size editing for local adjustments using feature-space composition, while the second enables a wide range of continuous style-element adjustments on the global level. For additional level-of-control, stroke-orientation adjustment is implemented utilizing the rotation-variance of CNNs when performing inference on rotated inputs. To achieve high output fidelity, we further add a patch-based style transfer method that enables users to obtain output resolutions of more than 20 Mpix. To enable an interactive editing experience, StyleTune uses Apple's CoreML and Metal APIs for ondevice inference on the mobile GPU featuring the dedicated Neural Engine. Our approach empowers users to create many novel results that are not possible with current mobile neural style transfer apps.

- Trezz, P., Pasewaldt, S., Klingbeil, M., Döllner, J., & Trapp, M. (2021). Forward Selfies. In Proceedings of the 21st ACM SIGGRAPH Appy Hour, ACM. DOI: 10.1145/3450415.3464400.

Abstract: Taking selfies is a common practice for smartphone users. Simultaneously capturing oneself and the desired background is not a trivial task, because it is often not possible to get a good view of both. Moreover, users often lose attention to their surroundings, thus taking a selfie also showed to lead to serious injuries. To ease the process of capturing selfies and to make it more safe, this work proposes forward selfies as a simple yet effective concept to account for both risk and challenges. Forward selfies seamlessly combine images of the front-facing and the rear-facing smartphone camera. We propose a mobile app that builds on this concept and implements the selfie synthesis in a post-processing image composition stage. Thereby, we can take advantage of the commonly more advanced back-camera hardware, i.e., providing higher image resolutions, larger field of views, and different perspectives. Finally, we leverage built-in camera optimizations for independently (de-)focusing objects at different distances, such as for persons and backgrounds. We conclude that the concept of forward selfies can effectively address and solve certain challenges of capturing selfies, which we demonstrate by a simple app user interface.

## 1.2.2 Arbeitspaket JD2 "KI-basierte Per-Frame Videoanalyse"

Im Mittelpunkt dieses Arbeitspakets standen die Anwendung von KI-Verfahren auf Bildsequenzen, die in Form digitaler Videos kodiert sind. Es ist hier nicht möglich, die Bilder (Frames) einzeln zu untersuchen bzw. zu transformieren, sondern sie müssen in ihrem zeitlichen Kontext analysiert werden, um zeitliche Kohärenz zu gewährleisten. Dazu ist pro Bild jeweils ein zeitlich relevanter Abschnitt in der Bildfolge zu identifizieren. Dieser Abschnitt bildet einen 2+1-dimensionalen Bildraum. Faktoren, die die Analyse beeinflussen, sind die Dynamik im ausgewählten zeitlichen Abschnitt und optional die synchron dazu gegebenen Audiodaten.

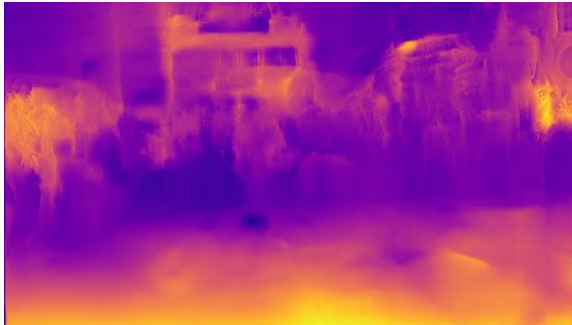
Es wurde ein Framework entwickelt, das es ermöglicht, verschiedene Analysen von Deep-Learning Methoden auf Videodaten anzuwenden. Dazu zählen insbesondere die Ableitung von einfachen (low-level) Angaben wie räumlicher Tiefeninformation und Oberflächenorientierung, optischer Fluss, sowie von Bildsegmentierungen. Weiterhin konnten unter Verwendung von DL-Methoden komplexe (high-level) Informationen wie beispielsweise die Detektion (Position und Anzahl) und Klassifikation von Objekten im Video abgeleitet werden. Darauf aufbauend wurden Ansätze untersucht und prototypisch implementiert, die es ermöglichen Dynamiken im Bild- und Audioströmen zu detektieren und klassifizieren. Diese Analyseergebnisse können u.a. verwendet werden um Videos automatisch zu verkürzen, unterschiedliche Stiltransformationen robust und kontrollierbar über den Verlauf des Video auf verschiedene Objekte anzuwenden. Die entwickelte Framework-Architektur ist modular und kann somit einfach erweitert werden, um schnell neue Verfahren und Techniken zu integrieren und evaluieren.



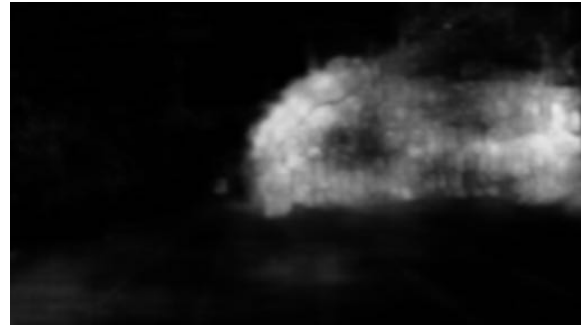
(a) Eingabeframe



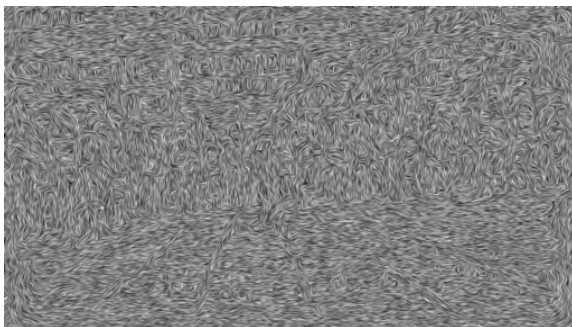
(b) Segmentierung und Kennzeichnung



(c) Tiefeninformation



(d) Visuelle Salienz



(e) Statischer Fluss im Frame



(f) Optischer Fluss zwischen Frames

Beispiele KI-basierter Per-Frame Videoanalyseergebnissen.

Für die praktischen Arbeiten, insbesondere zum Test des Frameworks und für die Demonstratoren, wurde hochspezialisierte Hardware und Speicherkapazität des KI-Labors herangezogen. Sehr erfolgreich konnten Trainings- und Testdaten dazu verwendet werden, um die neu entwickelten Verfahren für praxisnahe, realistische Szenarien testen zu können.

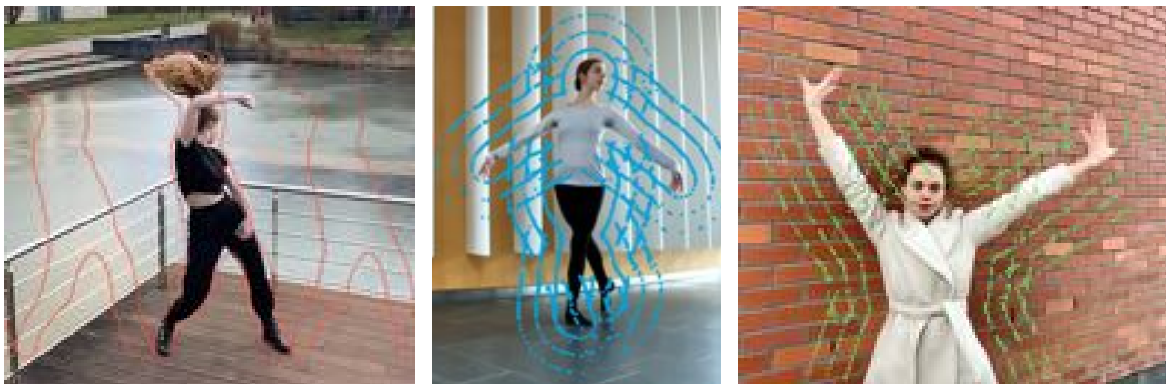
- Glöckner, D., Ihde, L., Döllner, J., & Trapp, M. (2020). Intermediate Representations for Vectorization of Stylized Images. *Journal of WSCG*, 28(1-2), 187-196.

Abstract: This paper presents a new approach for the vectorization of stylized images using intermediate data representations to interface image stylization and vectorization techniques. It enables the combination of efficient GPU-based implementations of interactive image stylization techniques and the advantages of vectorized image representations. We demonstrate the capabilities of our approach using half-toning and toon stylization techniques.

- Wagner, F., Döllner, J., & Trapp, M. (2020). Real-time Service-based Stream-processing of High-resolution Videos. *Journal of WSCG*, 28(1-2), 122-131.

Abstract: This paper reports on a service-based approach to enable real-time stream-processing of high-resolution videos. It presents a concept for integrating black-box image and video processing operations into a streaming framework. It further describes approaches to optimize data flow between the processing implementation and the framework to increase throughput and decrease latency. This enables the composition of streaming services to allow scaling for further throughput increase. We demonstrate the effectiveness of our approach by means of two real-time streaming-processing application examples.

Es wurden auf Grundlage eines professionellen Software-Frameworks zur digitalen Bild- und Video-Kodierung und -Dekodierung Deep-Learning-Methoden zur Analyse und Transformation von digitalen Videos weiterentwickelt und durch prototypischen Demonstratoren erprobt. Die entwickelten Techniken und Verfahren kamen hauptsächlich in folgenden Themenbereichen zur Erprobung: (1) der Visualisierung von menschlichen Bewegungen und (2) der Kontaktvisualisierung von menschlichen Gliedmaßen und Elementen der Umwelt.



*Abbildung: Beispiele für eine KI-basierte Visualisierung von Bewegungsinformationen durch per-Frame Analyse von User-generated Content Videos auf mobilen Endgeräten.*

Für die Visualisierung und Stilisierung von Bewegungsinformationen wurden ML-Verfahren zur Analyse, Ableitung, Verfeinerung und zeitlichen Stabilisierung von einfachen "Personen-Matten" (low-level) entwickelt und getestet. Auf dieser Basis wurden mit DL-Methoden komplexe (high-level) Informationen, wie beispielsweise die Detektion (Position und Anzahl) und Klassifikation von Objekten, im Video abgeleitet. Die Daten dienen als Eingabe für die Entwicklung von unterschiedlichen Techniken zur Bewegungsvisualisierung (siehe obiges Bild).

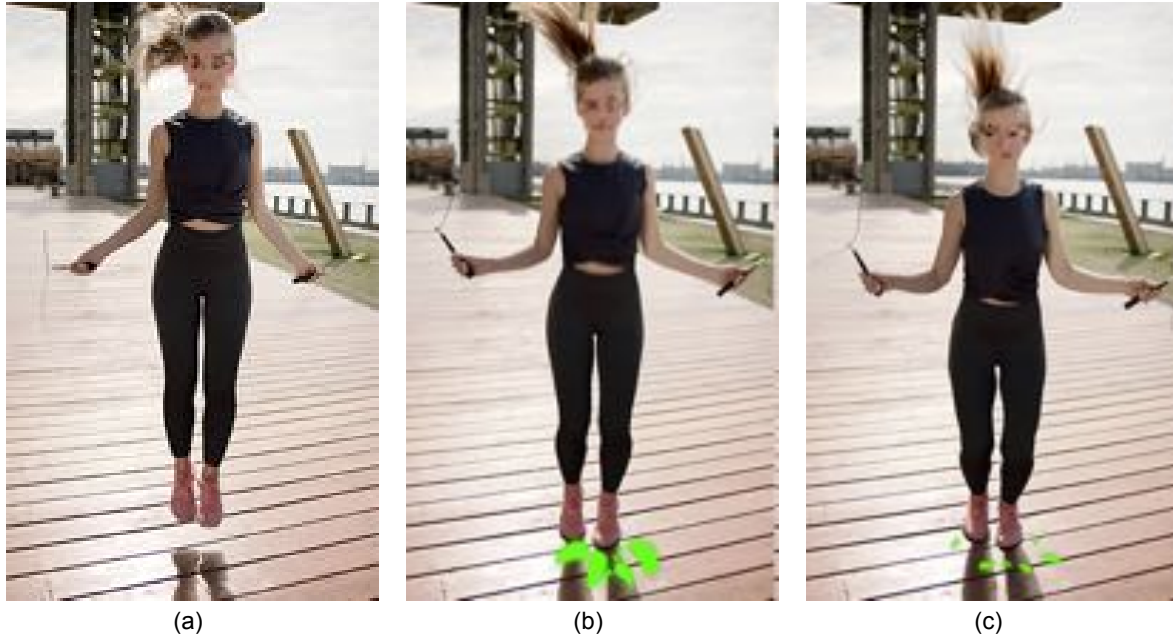


Abbildung: Beispielhafte Frames aus der entwickelten Kontaktvisualisierung für Videos: (a) Person, die den Boden nicht berührt, (b) eine Person, die den Boden berührt, ergibt einen starken Spritzeffekt, (c) nachfolgender Frame einer den Boden berührenden Person mit einem schwächeren Splash-Effekt.

Für die Visualisierung und Stilisierung von Kontaktinformationen (siehe obiges Bild) wurden ML-Verfahren zur Kontakterkennung erforscht, entwickelt und erprobt. Auf Basis eines weltweit einzigartigen Trainingsdatensatzes wurden zunächst DL-Netzwerke zur Ermittlung der Kontaktbereiche zwischen Mensch und Umwelt entwickelt. Die berechneten Informationen wurden anschließend von prototypischen Visualisierungstechniken verwendet.

Für die praktischen Arbeiten, insbesondere zum Test des Frameworks und für die Demonstratoren, wurde Hardware, Software und Speicherkapazität des KI-Labors verwendet. Sehr erfolgreich konnten Trainings- und Testdaten dazu verwendet werden, um die neu entwickelten Verfahren für praxisnahe, realistische Szenarien testen zu können. Die Bearbeitung des Arbeitspaketes wurde plangemäss durchgeführt; die wissenschaftlichen Ergebnisse wurden in den folgenden Publikationen zusammengefasst:

- Mayer, M., Trenz, P., Pasewaldt, S., Klingbeil, M., Döllner, J., Trapp, M., & Semmo, A. (2021). MotionViz – Artistic Visualization of Human Motion on Mobile Devices. In Proceedings of the 21st ACM SIGGRAPH Appy Hour ACM. DOI: 10.1145/3450415.3464398.

Abstract: We present MotionViz, an interactive iOS mobile app that enables users to amplify motion and dynamics in videos. MotionViz implements novel augmented reality and expressive rendering techniques in an end-to-end processing pipeline: multi-dimensional video data—including RGB, depth and segmentation information is captured, analyzed, and processed to render animated graphical elements that help express figures and actions. Through an easy to use graphical user interface, users can choose from a curated list of artistic motion visualization effects, including the overlay of animated silhouettes, halos, and contour lines. MotionViz is based on Apple's LiDAR technology, accelerated image processing APIs, and dedicated Neural Engine for real-time on-device processing.

- Reimann, M., Wegen, O., Pasewaldt, S., Semmo, A., Döllner, J., & Trapp, M. (2021). Teaching Data-driven Video Processing via Crowdsourced Data Collection. In Proceedings of the Eurographics Education Papers.

Abstract: This paper presents the concept and experience of teaching an undergraduate course on data-driven image and video processing. When designing visual effects that make use of Machine Learning (ML) models for image-based analysis or processing, the availability of training data typically represents a key limitation when it comes to feasibility and effect quality. The goal of our course is to enable students to implement new kinds of visual effects by acquiring training datasets via crowdsourcing that are used to train ML models as part of a video processing pipeline. First, we propose our course structure and best practices that are involved with crowdsourced data acquisitions. We then discuss the key insights we gathered from an exceptional undergraduate seminar project that tackles the challenging domain of video annotation and learning. In particular, we focus on how to practically develop annotation tools and collect high-quality datasets using Amazon Mechanical Turk (MTurk) in the budget- and time-constrained classroom environment. We observe that implementing the full acquisition and learning pipeline is entirely feasible for a seminar project, imparts hands-on problem solving skills, and promotes undergraduate research.

### **1.2.3 Arbeitspaket JD-3 "KI für Geospatial Analytics"**

3D-Punktwolken repräsentieren universell einsetzbare Geodaten, die durch eine Vielzahl von Geräten (z.B. Scanner, Drohnen, Laser) erfasst werden. Punktwolken bestehen aus atomar aufgeführten 3D-Punkten; sie sind ohne Semantik und Struktur. KI-Ansätze ermöglichen es, 3D-Punktwolken semantisch und strukturell zu interpretieren. Es wird dabei anhand des Skalenbereichs zwischen "Outdoor"- und "Indoor"-Punktwolken unterschieden. Für beide Kategorien sind als grundlegende Operationen Segmentierungs- und Klassifikationsverfahren auszuwählen, zu untersuchen und zu konfigurieren.

Die Nachfrage nach digitalen Indoor- und Outdoor-Zwillingen ist in den vergangenen Jahren in vielen Industriezweigen, z.B. Bauwesen, Infrastruktur-Monitoring und -Planung, stark gestiegen. Um digitale Zwillinge zu erstellen, werden LiDAR-Scanning und photogrammetrische Verfahren verwendet, welche die Umgebungen mit verschiedenen Detaillierungsgraden erfassen. Aus diesen Scans resultieren 3D-Punktwolken, welche als Hauptdatenquelle für die Erstellung von digitalen Zwillingen dienen. Ein zentrales Ergebnis unserer Arbeiten im KI-Labor für die KI-basierte Prozessierung von Indoor- und Outdoor-Punktwolken liegt in der Konstruktion eines "virtuellen 3D-Scanners":

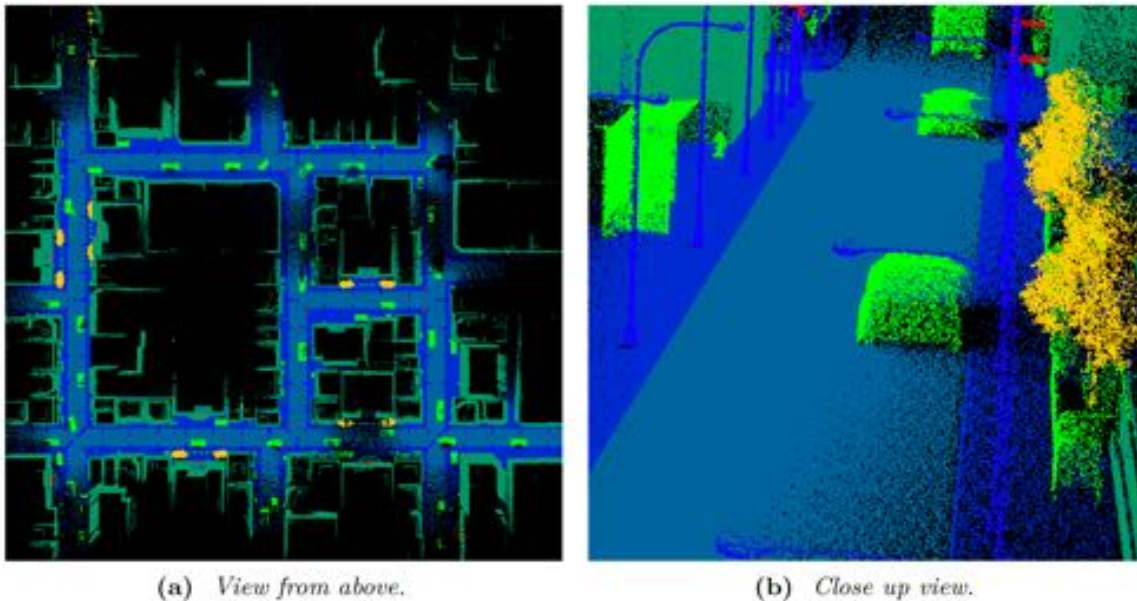


Abbildung: Synthetisch generierte 3D-Punktwolke zu einem gegebenen Gebiet innerhalb eines virtuellen 3D-Stadtmodells im Überblick (links) und aus einer 3D-Perspektive (rechts).

Im KI-Labor wurde eine neuartige Methode zur Generierung von klassifizierten synthetischen Punktwolken entwickelt und erprobt. Für eine effektive und anwendungsspezifische Nutzung müssen semantische Informationen aus den 3D-Punktwolken mittels Segmentierung und Klassifikation abgeleitet werden. Convolutional Neural Networks erlauben es, die Daten automatisch zu segmentieren und zu klassifizieren. Klassifizierte, repräsentative Daten werden benötigt, um solche Netzwerke zu trainieren. Das manuelle Klassifizieren ist sehr zeitaufwändig, weshalb andere Ansätze erforderlich sind. Virtuelle 3D-Modelle, wie z.B. hochdetaillierte 3D-Stadtmodelle, bilden die Ausgangsbasis: Ein virtueller 3D-Scanner fährt bzw. fliegt durch das 3D-Modell und tastet dabei die virtuelle 3D-Umgebung ab. Der Abtastprozess versucht dabei, die typischen Eigenheiten realer 3D-Scan-Prozesse zu imitieren, z.B. bzgl. ungleichmäßiger Dichte, Schwankung, Verschattung, Redundanz, etc. Konkret wird im vorgestellten Verfahren ein LiDAR-Gerät in einer virtuellen 3D-Umgebung simuliert, wobei durch das Abtasten der klassifizierten 3D-Meshoberflächen mit einem virtuellen Laserstrahl 3D-Punktwolken generiert werden. Das zugrunde liegende Ray-Tracing ermittelt die geometrische Position eines 3D-Punktes, verfügt aber aufgrund der 3D-Szenen-Modelle über semantische Information, die als Attribut dem 3D-Punkt beigefügt wird. Im Ergebnis liegen attributierte 3D-Punktwolken vor, die vollautomatisch aus 3D-Modellen synthetisiert werden und als validierte Trainingsdaten verwendet werden können. Die Ergebnisse der vLiDAR-Testdaten zeigen das enorme Potenzial des Ansatzes für die Generierung realistischerer Benchmark-Datensätze für die Punktwolkenanalyse.

- Döllner, J. (2020). Geospatial Artificial Intelligence: Potentials of Machine Learning for 3D Point Clouds and Geospatial Digital Twins. PFG Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88, 15–24. DOI: 10.1007/s41064-020-00102-3.

Abstract: Artificial Intelligence (AI) is changing fundamentally the way how IT solutions are built and operated across all application domains, including the geospatial domain. In this article, we briefly reflect on the term “AI” and outline the factors such as Machine Learning (ML) and Deep Learning (DL) that contribute to applying AI successfully for IT solutions. In the main part we discuss AI for the geospatial domain (GeoAI) focussing on 3D point clouds as a key category of geodata, describe their properties and discuss its suitability for ML and DL. In particular, we conclude that 3D point clouds constitute a corpus with similar properties than natural language corpora and formulate a naturalness hypothesis for 3D point clouds. We then outline concepts and examples of ML-based interpretation approaches that compute domain-specific and application-specific semantics for 3D point clouds without having to create explicit spatial models or explicit rule sets. Finally, we will show how ML enables us to efficiently build and maintain base data for digital twins of our environment such as virtual 3D city models, indoor models, or building information models.

- Stojanovic, V., Hagedorn, B., Trapp, M. & Döllner, J. (2020). Ontology-Driven Analytics for Indoor Point Clouds. In Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia.

Für die Durchführung der Arbeiten war insbesondere spezialisierte Hardware, Computing Services und lizenzierte Software-Frameworks des KI-Labors erforderlich; für die Testdaten sind umfangreiche Speicherkapazitäten (Storage Server, mobile Festplatten für den Datentransport und die gesicherte datenschutzkonforme Datenlagerung) erforderlich. Die erstellten Trainings- und Testdatensammlungen werden eingesetzt, um z.B. Präzision und Qualität von KI-basierten Analyseverfahren für 3D-Punktwolken zu messen.

#### **1.2.4 Arbeitspaket JD-4 "KI für Software Analytics"**

Für die Bewertung und Überwachung der Qualität eines Softwaresystems und der Qualität seines Entwicklungsprozesses ermöglichen KI-basierte Ansätze präzise und robuste Analysen. Software Repository Mining erlaubt es, insbesondere die Daten der Source-Code-Verwaltungs- und Ticketsysteme systematisch und vollautomatisch zu extrahieren. Diese Daten bilden die Ausgangsbasis für prädiktive Analysen zur Source-Code-Komplexität, zu Fehlerrisiken und zu langfristigen Aufwänden (“technische Schulden”). In diesem Arbeitspaket wurden ausgewählte Aspekte von KI-basierter Software Analytics untersucht. Im Gegensatz zu klassischen Ansätzen beruhen die entscheidenden Bewertungen auf zuvor – ggf. projektspezifisch – trainierten Daten, sodass keine expliziten regelbasierten oder entscheidungsbasierten Lösungskonzepte erforderlich sind, die extrem fehleranfällig und schwer handzuhaben wären.

Es wurden erste KI-basierte Verfahren konfiguriert, die mit Hilfe von Software Repository Mining Daten von Github auslesen und fusionieren. Die fusionierten Daten werden als hochdimensionale Merkmalsvektoren aufgefasst; sie bilden die Voraussetzung für die Ermittlung von Code-Evolution-Charakteristiken und Commit-Charakteristiken durch KI-basierte Verfahren, insbesondere Cluster- und Outlier-Detektion. Zur interaktiven Exploration dieser Merkmalsvektoren wurde ein grundlegendes, neuartiges Visualisierungsverfahren erforscht und erprobt.

- Wagner, L., Limberger, D., Scheibel, W., Trapp, M., & Döllner, J. (2020). A Framework for Interactive Exploration of Clusters in Massive Data Using 3D Scatter Plots and WebGL. In Proceedings of the 25th International Conference on 3D Web Technology, pp. 1-2. ACM. DOI: 10.1145/3424616.3424730.

Abstract: This paper presents a rendering framework for the visualization of massive point datasets in the web. It includes highly interactive point rendering, cluster visualization, basic interaction methods, and importance-based labeling, while being available for both mobile and desktop browsers. The rendering style is customizable [...]. Our evaluation indicates that the framework facilitates interactive visualization of tens of millions of raw data points even without dynamic filtering or aggregation.

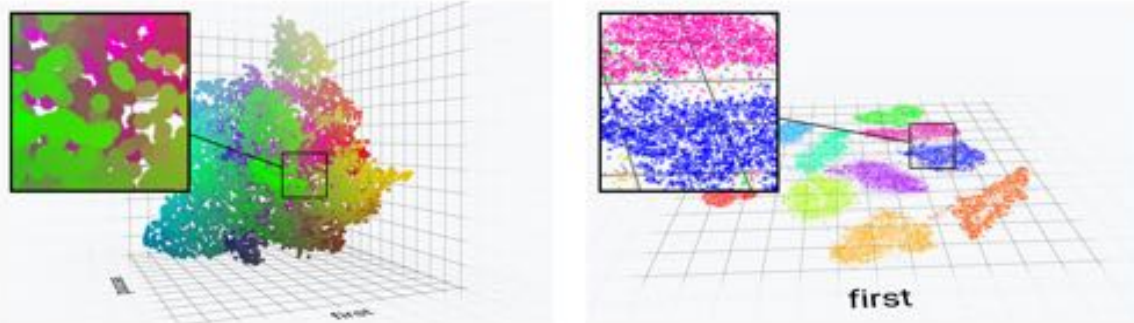
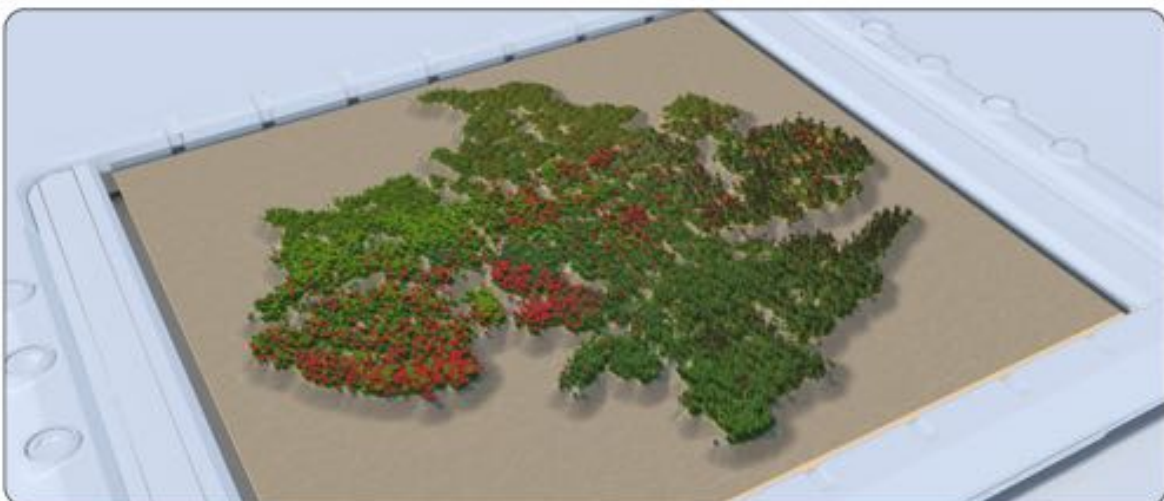


Abbildung: 3D-Scatter-Plotter zur Visualisierung von hochdimensionalen Merkmalsvektoren.

Es wurde darüber hinaus ein neuartiges Verfahren für die Visualisierung typischer KI-basierter Analyseergebnisse für Code- und Commit-Daten aus Software Repositorien erforscht:

- Atzberger, D., Cech, T., de la Haye, M., Söchting, M., Scheibel, W., Limberger, D. and Döllner, J. (2021). Software Forest: A Visualization of Semantic Similarities in Source Code using a Tree Metaphor. In Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 112-122. Scitepress. DOI: 10.5220/0010267601120122.

Abstract: Software visualization techniques provide effective means for program comprehension tasks as they allow developers to interactively explore large code bases. A frequently encountered task during software development is the detection of source code files of similar semantics. To assist this task we present Software Forest, a novel 2.5D software visualization that enables interactive exploration of semantic similarities within a software system, illustrated as a forest. The underlying layout results from the analysis of the vocabulary of the software documents using Latent Dirichlet Allocation and Multidimensional Scaling and therefore reflects the semantic similarity between source code files. By mapping properties of a software entity, e.g., size metrics or trend data, to visual variables encoded by various, figurative tree meshes, aspects of a software system can be displayed. This concept is complemented with implementation details as well as a discussion on applications.



*Abbildung: Beispiel eines "Software Forest", resultierend aus Dimensionsreduktion und KI-basierter Berechnung hochdimensionaler Merkmalsvektoren zu Code Units eines Softwaresystems, wobei semantische Ähnlichkeit layoutbildend eingesetzt wird; dominante Topics konstituieren jeweils eine eigene Baumart.*

In den Arbeitspaketen JD-1 bis JD-4 finden sich softwaretechnisch jeweils massiv parallele Implementierungen mit computergrafischen und geometrischen Algorithmen, die spezialisierte Labor-Hardware, Labor-Software, Test- und Trainingsdatensammlungen und Speicherkapazitäten des KI-Labors erforderten.

## 1.3 Arbeitsgruppe Prof. Tobias Friedrich

### 1.3.1 Arbeitspaket TF1 "Testdatensätze für SAT"

Das Erfüllbarkeitsproblem der Aussagenlogik (SAT) ist das klassische Problem künstlichen Intelligenz. Viele andere Optimierungs- und Entscheidungsproblem lassen sich in SAT ausdrücken und lösen. Typische Beispiele sind Scheduling, Model Checking und VLSI-Design. In der Theoretischen Informatik ist seit den 70er-Jahren bekannt, dass manche SAT-Instanzen beweisbar nur sehr schwer lösbar sind. Dennoch sind heutzutage industrielle SAT-Solver in der Lage, Instanzen mit Millionen von Variablen in kurzer Zeit zu lösen. Für eine Weiterentwicklung der SAT-Solver bedarf es realistischer, großer Testdatensätze. Diese sind aufgrund der industriellen Bedeutung nicht allgemein verfügbar. In diesem Arbeitspaket sollen derartige Instanzen gesammelt und modelliert werden, um automatisch große realistische SAT-Instanzen erzeugen zu können. Diese Fragestellung gilt als eine der "10 großen Fragen der SAT-Forschung".

Das Arbeitspaket wurde plangemäss durchgeführt, in dem Instanzen gesammelt und modelliert wurden. Die bisherigen wissenschaftlichen Ergebnisse werden wie folgt zusammengefasst:

- Bläsius, T., Friedrich, T., Göbel, A., Levy, J., & Rothenberger, R. (2021). The Impact of Heterogeneity and Geometry on the Proof Complexity of Random Satisfiability. In Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, pp. 42-53. ACM-SIAM. DOI: 10.1137/1.9781611976465.4.

Abstract: Boolean satisfiability is one of the most studied basic AI problems. The hardness of Boolean formulas at the solubility phase transition of random propositional satisfiability (SAT) has been intensely studied for decades both empirically and theoretically. Solvers based on stochastic local search (SLS) appear to scale very well at the critical threshold, while complete backtracking solvers exhibit exponential scaling. On industrial SAT instances, this phenomenon is inverted: backtracking solvers can tackle large industrial problems, where SLS-based solvers appear to stall. Industrial instances exhibit sharply different structure than uniform random instances. Among many other properties, they are often heterogeneous in the sense that some variables appear in many while others appear in only few clauses. In our recent work "On the Empirical Time Complexity of Scale-Free 3-SAT at the Phase Transition" (TACAS 19) we conjecture that the heterogeneity of SAT formulas alone already contributes to the trade-off in performance between SLS solvers and complete backtracking solvers. We empirically determine how the run time of SLS vs. backtracking solvers depends on the heterogeneity of the input, which is controlled by drawing variables according to a scale-free distribution. Our experiments reveal that the efficiency of complete solvers at the phase transition is strongly related to the heterogeneity of the degree distribution. We report results that suggest the depth of satisfying assignments in complete search trees is influenced by the level of heterogeneity as measured by a

power-law exponent. We also find that incomplete SLS solvers, which scale well on uniform instances, are not affected by heterogeneity. The main contribution of this paper utilizes the scale-free random 3-SAT model to isolate heterogeneity as an important factor in the scaling discrepancy between complete and SLS solvers at the uniform phase transition found in previous works.

More recently and still unpublished, we have studied the unexplained gap between theory and practice in SAT research (available at <https://arxiv.org/abs/2004.07319>): Satisfiability is considered the canonical NP-complete problem and is used as a starting point for hardness reductions in theory, while in practice heuristic SAT solving algorithms can solve large-scale industrial SAT instances very efficiently. This disparity between theory and practice is believed to be a result of inherent properties of industrial SAT instances that make them tractable. Two characteristic properties seem to be prevalent in the majority of real-world SAT instances, heterogeneous degree distribution and locality. To understand the impact of these two properties on SAT, we study the proof complexity of random  $k$ -SAT models that allow to control heterogeneity and locality. Our findings show that heterogeneity alone does not make SAT easy as heterogeneous random  $k$ -SAT instances have superpolynomial resolution size. This implies intractability of these instances for modern SAT-solvers. On the other hand, modeling locality with an underlying geometry leads to small unsatisfiable subformulas, which can be found within polynomial time.

A key ingredient for the result on geometric random  $k$ -SAT can be found in the complexity of higher-order Voronoi diagrams. As an additional technical contribution, we show a linear upper bound on the number of non-empty Voronoi regions, that holds for points with random positions in a very general setting. In particular, it covers arbitrary  $p$ -norms, higher dimensions, and weights affecting the area of influence of each point multiplicatively. This is in stark contrast to quadratic lower bounds for the worst case.

- Casel, K., Dreier, J., Fernau, H., Gobbert, M., Kuinke, P., Villaamil, F., Schmid, M., & van Leeuwen, E. (2020). Complexity of independency and cliquy trees. *Discrete Applied Mathematics*, 272(2-15), 2-15. DOI: 10.1016/j.dam.2018.08.011.

Abstract: An independency (cliquy) tree of an  $n$ -vertex graph  $G$  is a spanning tree of  $G$  in which the set of leaves induces an independent set (clique). We study the problems of minimizing or maximizing the number of leaves of such trees, and fully characterize their parameterized complexity. We show that all four variants of deciding if an independency/cliquy tree with at least/most  $l$  leaves exists parameterized by  $l$  are either Para-NP-or  $W[1]$ -hard. We prove that minimizing the number of leaves of a cliquy tree parameterized by the number of internal vertices is Para-NP-hard too. However, we show that minimizing the number of leaves of an independency tree parameterized by the number  $k$  of internal vertices has an  $O^*(4^k)$ -time algorithm and a  $2k$  vertex kernel. Moreover, we prove that maximizing the number  $k$  of internal vertices both have an  $O^*(18^k)$ -time algorithm and an  $O(k2^k)$  vertex kernel, but no polynomial kernel unless the polynomial hierarchy collapses to the third level. Finally, we present an  $O(3^{nf(n)})$ -time algorithm to find a spanning tree where the leaf set has a property that can be decided in  $f(n)$  time and has minimum or maximum size.

Das Arbeitspaket wurde plangemäss durchgeführt, indem Instanzen gesammelt und modelliert wurden. Die wissenschaftlichen Ergebnisse werden wie folgt zusammengefasst:

- Bläsius, T., Friedrich, T., Göbel, A., Levy, J., & Rothenberger, R. (2021). The Impact of Heterogeneity and Geometry on the Proof Complexity of Random Satisfiability. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms*, pp. 42-53. DOI: 10.1137/1.9781611976465.4.

Abstract: More recently and just published at SODA'21, we have studied the unexplained gap between theory and practice in SAT research: Satisfiability is considered the canonical NP-complete problem and is used as a starting point for hardness reductions in theory, while in practice heuristic SAT solving algorithms can solve large-scale industrial SAT instances very efficiently. This disparity between theory and practice is believed to be a result of inherent properties of industrial SAT instances that make them tractable. Two characteristic properties seem to be prevalent in the majority of real-world SAT instances, heterogeneous degree distribution and locality. To understand the impact of these two properties on SAT, we study the proof complexity of random  $k$ -SAT models that allow to control heterogeneity and locality. Our findings show that heterogeneity alone does not make SAT easy as heterogeneous random  $k$ -SAT instances have superpolynomial resolution size. This implies intractability of these instances for modern SAT-solvers. On the other hand, modeling locality with an underlying geometry leads to small

unsatisfiable subformulas, which can be found within polynomial time.

A key ingredient for the result on geometric random k-SAT can be found in the complexity of higher-order Voronoi diagrams. As an additional technical contribution, we show a linear upper bound on the number of non-empty Voronoi regions, that holds for points with random positions in a very general setting. In particular, it covers arbitrary p-norms, higher dimensions, and weights affecting the area of influence of each point multiplicatively. This is in stark contrast to quadratic lower bounds for the worst case.

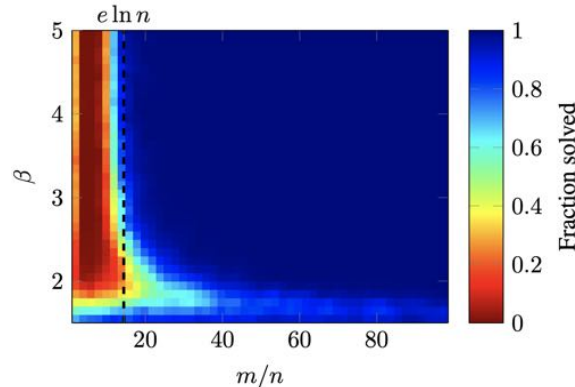


Abbildung: Untersuchung der Erfüllbarkeit von skalenfreien 3-SAT Formeln. Dargestellt ist der Anteil der Formeln, die der Greedy Algorithmus von Koutsoupias-Papadimitriou (1992) löst, abhängig von der Constraint Dichte  $m/n$  und dem Power-Law-Exponent  $\beta$ .

- Casel, K., Dreier, J., Fernau, H., Gobbert, M., Kuinke, P., Villaamil, F., Schmid, M., & van Leeuwen, E. (2020). Complexity of independency and cliquy trees. *Discrete Applied Mathematics*, 272(2-15), 2-15. DOI: 10.1016/j.dam.2018.08.011.

Abstract: An independency (cliquy) tree of an  $n$ -vertex graph  $G$  is a spanning tree of  $G$  in which the set of leaves induces an independent set (clique). We study the problems of minimizing or maximizing the number of leaves of such trees, and fully characterize their parameterized complexity. We show that all four variants of deciding if an independency/cliquy tree with at least/most  $l$  leaves exists parameterized by  $l$  are either Para-NP- or  $W[1]$ -hard. We prove that minimizing the number of leaves of a cliquy tree parameterized by the number of internal vertices is Para-NP-hard too. However, we show that minimizing the number of leaves of an independency tree parameterized by the number  $k$  of internal vertices has an  $O^*(4^k)$ -time algorithm and a  $2k$  vertex kernel. Moreover, we prove that maximizing the number  $k$  of internal vertices both have an  $O^*(18^k)$ -time algorithm and an  $O(k2^k)$  vertex kernel, but no polynomial kernel unless the polynomial hierarchy collapses to the third level. Finally, we present an  $O(3^{nf(n)})$ -time algorithm to find a spanning tree where the leaf set has a property that can be decided in  $f(n)$  time and has minimum or maximum size.

### 1.3.2 Arbeitspaket TF2 "Testdatensätze für Netzwerke"

Effiziente Algorithmen bilden den Kern komplexer KI-Systeme. Die Performance der einzelnen Algorithmen ist entscheidend für die Gesamt-Performance des Systems. Daher ist die Untersuchung und Beschleunigung dieser Basisalgorithmen von entscheidender Bedeutung für KI-Systeme. Wie sich innerhalb der Algorithmik in den letzten zehn Jahren herausgestellt hat, ist die häufigste Datenart, die ein Algorithmus verarbeitet, ein Netzwerk. Dieses kann je nach Algorithmus strukturell sehr verschieden sein – beispielsweise gerichtet oder ungerichtet, Knoten- oder Kanten-gewichtet, mit oder ohne Geometrie, oder basierend auf sehr verschiedenen Topologien. Ein umfangreicher Netzwerk-Testdatensatz unterstützt daher die Weiterentwicklung zahlreicher Basis-Algorithmen. In diesem Arbeitspaket wurden vorhandene realistische Netzwerk-Testdatensätze gesammelt und modelliert. Die

wichtigsten Graph-Parametern wurden für jeden Netzwerk-Datensatz bestimmt und dokumentiert.

Das Arbeitspaket wurde plangemäss durchgeführt, in dem Instanzen gesammelt und modelliert wurden. Die wissenschaftlichen Ergebnisse werden wie folgt zusammengefasst:

- Cohen, S., Hershcovitch, M., Taraz, M.; Kißig, O.; Wood, A.; Waddington, D.; Chin, P., & Friedrich, T. (2022). Drug Repurposing using Link Prediction on Knowledge Graphs with Applications to Non-Volatile Memory. In Proceedings of the International Conference on Complex Networks and Their Applications, pp. 742-753. Springer.

Abstract: The active global SARS-CoV-2 pandemic caused more than 167 million cases and 3.4 million deaths worldwide. The development of completely new drugs for such a novel disease is a challenging, time intensive process and despite researchers around the world working on this task, no effective treatments have been developed yet. This emphasizes the importance of drug repurposing, where treatments are found among existing drugs that are meant for different diseases. A common approach to this is based on knowledge graphs, that condense relationships between entities like drugs, diseases and genes. Graph neural networks (GNNs) can then be used for the task at hand by predicting links in such knowledge graphs. Expanding on state-of-the-art GNN research, Doshi et al. recently developed the Dr-COVID model. We further extend their work using additional output interpretation strategies. The best aggregation strategy derives a top-100 ranking of candidate drugs, 32 of which currently being in COVID-19-related clinical trials. Moreover, we present an alternative application for the model, the generation of additional candidates based on a given pre-selection of drug candidates using collaborative filtering. In addition, we improved the implementation of the Dr-COVID model by significantly shortening the inference and pre-processing time by exploiting data-parallelism. As drug repurposing is a task that requires high computation and memory resources, we further accelerate the post-processing phase using a new emerging hardware --- we propose a new approach to leverage the use of high-capacity Non-Volatile Memory for aggregate drug ranking.

- Krogmann, S., Lenzner, P., Molitor, L., & Skopalik, A. (2021). Two-Stage Facility Location Games with Strategic Clients and Facilities. In Proceedings of the 30th International Joint Conference on Artificial Intelligence, pp. 292-298. DOI: 10.24963/ijcai.2021/41.

Abstract: We consider non-cooperative facility location games where both facilities and clients act strategically and heavily influence each other. This contrasts established game-theoretic facility location models with non-strategic clients that simply select the closest opened facility. In our model, every facility location has a set of attracted clients and each client has a set of shopping locations and a weight that corresponds to her spending capacity. Facility agents selfishly select a location for opening their facility to maximize the attracted total spending capacity, whereas clients strategically decide how to distribute their spending capacity among the opened facilities in their shopping range. We focus on a natural client behavior similar to classical load balancing: our selfish clients aim for a distribution that minimizes their maximum waiting times for getting serviced, where a facility's waiting time corresponds to its total attracted client weight. We show that subgame perfect equilibria exist and give almost tight constant bounds on the Price of Anarchy and the Price of Stability, which even hold for a broader class of games with arbitrary client behavior. Since facilities and clients influence each other, it is crucial for the facilities to anticipate the selfish clients' behavior when selecting their location. For this, we provide an efficient algorithm that also implies an efficient check for equilibrium. Finally, we show that computing a socially optimal facility placement is NP-hard and that this result holds for all feasible client weight distributions.

- Kißig, O., Taraz, M., Cohen, S., Doskoč, V., & Friedrich, T.. Drug Repurposing for Multiple COVID Strains using Collaborative Filtering. In Proceedings of the ICLR Workshop on Machine Learning for Preventing and Combating Pandemics.

Abstract: The ongoing COVID-19 pandemic demands for a swift discovery of suitable treatments. The development of completely new compounds for such a novel disease is a challenging, time intensive process. This amplifies the relevance of drug repurposing, a technique where existing drugs are used to treat other diseases. A common bioinformatical approach to this is based on knowledge graphs, which compile relationships between drugs,

diseases, genes and other biomedical entities. Then, graph neural networks (GNNs) are used for the drug repurposing task as they provide a good link prediction performance on such knowledge graphs. Building on state-of-the-art GNN research, Doshi & Chepuri (2020) construct the remarkable model DR-COVID. We re-implement their model and extend the approach to perform significantly better. We propose and evaluate several strategies for the aggregation of link predictions into drug recommendation rankings. With the help of clustering of similar target diseases we improve the model by a substantial margin, compiling a top-100 ranking of candidates including 32 currently being in COVID-19-related clinical trials. Regarding the re-implementation, we offer more flexibility in the selection of the graph neighborhood sizes fed into the model and reduce the training time significantly by making use of data parallelism.

- Friedrich, T., Neumann, F., Rothenberger, R., & Sutton, A. (2021). Solving Non-Uniform Planted and Filtered Random SAT Formulas Greedily. In Proceedings of the International Conference on Theory and Applications of Satisfiability Testing, pp. 188-206. Springer.

Abstract: Recently, there has been an interest in studying non-uniform random  $k$ -satisfiability ( $k$ -SAT) models in order to address the non-uniformity of formulas arising from real-world applications. While uniform random  $k$ -SAT has been extensively studied from both a theoretical and experimental perspective, understanding the algorithmic complexity of heterogeneous distributions is still an open challenge. When a sufficiently dense formula is guaranteed to be satisfiable by conditioning or a planted assignment, it is well-known that uniform random  $k$ -SAT is easy on average. We generalize this result to the broad class of non-uniform random  $k$ -SAT models that are characterized only by an ensemble of distributions over variables with a mild balancing condition. This balancing condition rules out extremely skewed distributions in which nearly half the variables occur less frequently than a small constant fraction of the most frequent variables, but generalizes recently studied non-uniform  $k$ -SAT distributions such as power-law and geometric formulas. We show that for all formulas generated from this model of at least logarithmic densities, a simple greedy algorithm can find a solution with high probability. As a side result we show that the total variation distance between planted and filtered (conditioned on satisfiability) models is  $o(1)$  once the planted model produces formulas with a unique solution with probability  $1-o(1)$ . This holds for all random  $k$ -SAT models where the signs of variables are drawn uniformly and independently at random.

- Bläsius, T., Friedrich, T., Göbel, A., Levy, J., & Rothenberger, R. (2021). The Impact of Heterogeneity and Geometry on the Proof Complexity of Random Satisfiability. In Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, pp. 42-53. DOI: 10.1137/1.9781611976465.4.

Abstract: More recently and just published at SODA'21, we have studied the unexplained gap between theory and practice in SAT research: Satisfiability is considered the canonical NP-complete problem and is used as a starting point for hardness reductions in theory, while in practice heuristic SAT solving algorithms can solve large-scale industrial SAT instances very efficiently. This disparity between theory and practice is believed to be a result of inherent properties of industrial SAT instances that make them tractable. Two characteristic properties seem to be prevalent in the majority of real-world SAT instances, heterogeneous degree distribution and locality. To understand the impact of these two properties on SAT, we study the proof complexity of random  $k$ -SAT models that allow to control heterogeneity and locality. Our findings show that heterogeneity alone does not make SAT easy as heterogeneous random  $k$ -SAT instances have superpolynomial resolution size. This implies intractability of these instances for modern SAT-solvers. On the other hand, modeling locality with an underlying geometry leads to small unsatisfiable subformulas, which can be found within polynomial time.

A key ingredient for the result on geometric random  $k$ -SAT can be found in the complexity of higher-order Voronoi diagrams. As an additional technical contribution, we show a linear upper bound on the number of non-empty Voronoi regions, that holds for points with random positions in a very general setting. In particular, it covers arbitrary  $p$ -norms, higher dimensions, and weights affecting the area of influence of each point multiplicatively. This is in stark contrast to quadratic lower bounds for the worst case.

### 1.3.3 Arbeitspaket TF-3 "Niedrigdimensionale Einbettung"

Viele Daten in KI-Systemen sind sehr hochdimensional und damit nicht nur für den Menschen, sondern auch für viele Algorithmen schwer zu greifen und zu verarbeiten. Die

intrinsische Komplexität hochdimensionaler Daten ist oftmals weit geringer und basiert auf einer niedrigdimensionalen Darstellung. Auch beispielsweise Netzwerke (vgl. TF-2) können als hochdimensionale Objekte angesehen werden. Ein Netzwerk kann eine scheinbar-komplexe kombinatorische Struktur aufzeigen, auch wenn diese sich direkt aus der Nachbarschaftsbeziehung in einer zwei-dimensionalen euklidischen Ebene herleiten lässt. Auch andere reale kombinatorische Strukturen lassen sich durch eine niedrigdimensionale Einbettung leichter darstellen. Eine solche niedrigdimensionale Einbettung erlaubt schnellere Basisalgorithmen für KI-Systeme.

Das Arbeitspaket wurde plangemäß durchgeführt und ein neuer Algorithmus entwickelt, um effektiv und effizient große skalenfreie Netzwerke in die hyperbolische Ebene einzubetten:

- Bläsius, T., Friedrich, T., & Katzmann, M. (2021). Force-Directed Embedding of Scale-Free Networks in the Hyperbolic Plane. In Proceedings of the Symposium on Experimental Algorithms. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. DOI: 10.4230/LIPIcs.SEA.2021.22.

Abstract: Force-directed drawing algorithms are the most commonly used approach to visualize networks. While they are usually very robust, the performance of Euclidean spring embedders decreases if the graph exhibits the high level of heterogeneity that typically occurs in scale-free real-world networks. As heterogeneity naturally emerges from hyperbolic geometry (in fact, scale-free networks are often perceived to have an underlying hyperbolic geometry), it is natural to embed them into the hyperbolic plane instead. Previous techniques that produce hyperbolic embeddings usually make assumptions about the given network, which (if not met) impairs the quality of the embedding. It is still an open problem to adapt force-directed embedding algorithms to make use of the heterogeneity of the hyperbolic plane, while also preserving their robustness. We identify fundamental differences between the behavior of spring embedders in Euclidean and hyperbolic space, and adapt the technique to take advantage of the heterogeneity of the hyperbolic plane.

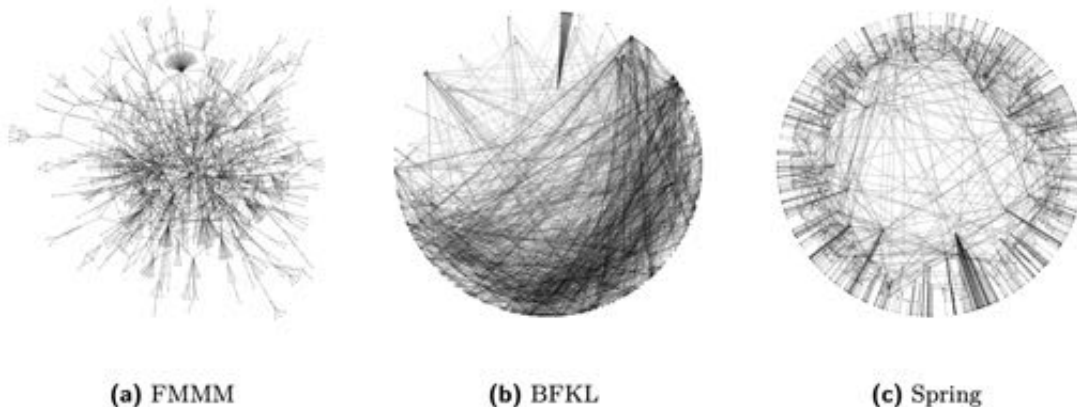


Abbildung: Drei verschiedene hyperbolische Einbettungen der größten Komponente (1458 Knoten) des Netzwerkes "bio-yeast-protein-inter".

Ein weiteres klassisches KI-Problem, welches als niedrig-dimensionale Einbettung beschrieben werden kann, ist Clustering. In mehreren unveröffentlichten Arbeiten haben wir uns mit Clustering beschäftigt. Dabei wurde insbesondere eine neue Struktur "balanced crown decomposition" eingeführt, um zusammenhängende Teilgraphen effizienter identifizieren zu können. Dies hat Anwendungen im Bereich der Kernelisierung und für Approximationsalgorithmen.

- Casel, K., Friedrich, T., Isaac, D., Niklanovits, A., & Zeif, Z. (2021). Balanced Crown Decomposition for Connectivity Constraints, In Proceedings of the 29th Annual European Symposium on Algorithms, 26, 1-26. DOI: 10.4230/LIPIcs.ESA.2021.26.

Abstract: We introduce the balanced crown decomposition that captures the structure imposed on graphs by their connected induced subgraphs of a given size. Such subgraphs are a popular modeling tool in various application areas, where the non-local nature of the connectivity condition usually results in very challenging algorithmic tasks. The balanced crown decomposition is a combination of a crown decomposition and a balanced partition which makes it applicable to graph editing as well as graph packing and partitioning problems. We illustrate this by deriving improved approximation algorithms and kernelization for a variety of such problems. In particular, through this structure, we obtain the first constant-factor approximation for the Balanced Connected Partition (BCP) problem, where the task is to partition a vertex-weighted graph into  $k$  connected components of approximately equal weight. We derive a 3-approximation for the two most commonly used objectives of maximizing the weight of the lightest component or minimizing the weight of the heaviest component.

### 1.3.4 Arbeitspaket TF-4 "Automatisches Parameter-Tuning"

Während der Entwicklung von komplexen Softwaresystemen werden kontinuierlich Designentscheidungen getroffen, die das resultierende Verhalten der Software entscheidend beeinflussen. Es gibt nicht den einen optimalen Ansatz, der für jede zukünftige Verwendung der Software der richtige ist. Je nach Eingabe, ist ein anderer Algorithmus oder eine andere Parameterwahl günstiger. Diese Information liegt zum Zeitpunkt der Implementierung des Softwaresystems jedoch nicht vor. Gerade bei besonders rechenaufwändigen KI-Systemen gibt es keinen etablierten State-of-the-art Algorithmus, sondern am effizientesten ist je nach Eingabe eine Sammlung ganz verschiedener Algorithmen mit komplementären Eigenschaften. Diese Komplementarität wird in diesem Arbeitspaket genutzt, um für jede Eingabeinstanz automatisch denjenigen Algorithmus und diejenige Parametrisierung heraus zu suchen, welcher die beste erwartete Performance liefert.

Das Arbeitspaket wurde plangemäß durchgeführt anhand des klassischen Hitting Set Problems. Die beiden HPI-Fachgebiete Algorithm Engineering und Information Systems haben gemeinsam einen parametrisierten Algorithmus entwickelt zum Aufzählen von Unique Column Combinations, einem wichtigen Problem in der Analyse von großen Datenbanken.

- Birnick, J., Bläsius, T., Friedrich, T., Naumann, F., Papenbrock, T., & Schirneck, M. (2020). Hitting set enumeration with partial information for unique column combination discovery. In Proceedings of the VLDB Endowment, 13(12), pp. 2270-2283. DOI: 10.14778/3407790.3407824.

Abstract: Unique column combinations (UCCs) are a fundamental concept in relational databases. They identify entities in the data and support various data management activities. Still, UCCs are usually not explicitly defined and need to be discovered. State-of-the-art data profiling algorithms are able to efficiently discover UCCs in moderately sized datasets, but they tend to fail on large and, in particular, on wide datasets due to run time and memory limitations.

In this paper, we introduce HPIValid, a novel UCC discovery algorithm that implements a faster and more resource-saving search strategy. HPIValid models the metadata discovery as a hitting set enumeration problem in hypergraphs. In this way, it combines efficient discovery techniques from data profiling research with the most recent theoretical insights into enumeration algorithms. Our evaluation shows that HPIValid is not only orders of magnitude faster than related work, it also has a much smaller memory footprint.

Viele Probleme des maschinellen Lernens lassen sich als submodulare Funktion beschreiben. Das effiziente Maximieren von submodularen Funktionen ist daher

Basisproblem mit zahlreichen Parametern. Die folgende Arbeit schlägt ein neues adaptives Sampling-Verfahren hierfür vor.

- Quinzan, F., Doskoč, V., Göbel, A., & Friedrich, T. (2021). Adaptive Sampling for Fast Constrained Maximization of Submodular Functions. In Proceedings of the 24th International Conference on Artificial Intelligence and Statistics, pp. 964-972. PMLR.

Abstract: Several large-scale machine learning tasks, such as data summarization, can be approached by maximizing functions that satisfy submodularity. These optimization problems often involve complex side constraints, imposed by the underlying application. In this paper, we develop an algorithm with poly-logarithmic adaptivity for non-monotone submodular maximization under general side constraints. The adaptive complexity of a problem is the minimal number of sequential rounds required to achieve the objective. Our algorithm is suitable to maximize a non-monotone submodular function under a  $p$ -system side constraint, and it achieves a  $(p+O(\sqrt{p}))$ -approximation for this problem, after only poly-logarithmic adaptive rounds and polynomial queries to the valuation oracle function. Furthermore, our algorithm achieves a  $(p+O(1))$ -approximation when the given side constraint is a  $p$ -extendible system.

This algorithm yields an exponential speed-up, with respect to the adaptivity, over any other known constant-factor approximation algorithm for this problem. It also competes with previous known results in terms of the query complexity. We perform various experiments on various real-world applications. We find that, in comparison with commonly used heuristics, our algorithm performs better on these instances.

- Angrick, S., Bals, B., Hastrich, N., Kleissl, M., Schmidt, J., Doskoč, V., Katzmann, M., Molitor, L., & Friedrich, T. (2022). Towards Explainable Real Estate Valuation via Evolutionary Algorithms. In Proceedings of the Genetic and Evolutionary Computation Conference, pp. 1130-1138. ACM. DOI: 10.1145/3512290.3528801.

Abstract: Human lives are increasingly influenced by algorithms, which therefore need to meet higher standards not only in accuracy but also with respect to explainability. This is especially true for high-stakes areas such as real estate valuation. Unfortunately, the methods applied there often exhibit a trade-off between accuracy and explainability. One explainable approach is *case-based reasoning (CBR)*, where each decision is supported by specific previous cases. However, such methods can be wanting in accuracy. The unexplainable machine learning approaches are often observed to provide higher accuracy but are not scrutible in their decision-making. One explainable approach is *case-based reasoning (CBR)*, where each decision is supported by specific previous cases. However, such methods can be wanting in accuracy. The unexplainable machine learning approaches are often observed to provide higher accuracy but are not scrutible in their decision-making. In this paper, we apply *evolutionary algorithms (EAs)* to CBR predictors in order to improve their performance. In particular, we deploy EAs to the similarity functions (used in CBR to find comparable cases), which are fitted to the data set at hand. As a consequence, we achieve higher accuracy than state-of-the-art deep neural networks (DNNs), while keeping interpretability and explainability. These results stem from our empirical evaluation on a large data set of real estate offers where we compare known similarity functions, their EA-improved counterparts, and DNNs. Surprisingly, DNNs are only on par with standard CBR techniques. However, using EA-learned similarity functions does yield an improved performance.

## 1.4 Arbeitsgruppe Prof. Robert Hirschfeld

### 1.4.1 Arbeitspaket RH1 "KI für Predictive-Software-Maintenance"

Um Entwickler bei der Wartung von komplexem Programm-Code zu unterstützen, können KI-basierte Verfahren eingesetzt werden. Insbesondere die Analyse von Software-Repositories hinsichtlich der Verteilung und Interaktion problemraumspezifischer Konzepte vereinfacht das Programmverständnis. Die Fähigkeit, formale Programmiersprachen mit natürlichsprachlichen Artefakten (z.B. Dokumentation, Benennung

von Programmelementen, oder Test- und Laufzeitdaten) in Bezug zu setzen, ist eine der vielversprechendsten Eigenschaften KI-basierter Verfahren. Künftige Forschung soll darüber hinausgehend konkrete Vorschläge zur Verbesserung der Programm- verständlichkeit, -änderbarkeit und -testbarkeit generieren. Von großem Interesse für die KI-basierte Entwicklung von IT-Systemtechnik und dem Software Engineering sind hierbei auch aktuelle Datensätze, anhand derer die Vorhersagekraft neuer Verfahren vergleichbar gemessen werden kann.

Das Arbeitspaket wurde plangemäss durchgeführt, indem Instanzen gesammelt und modelliert wurden. Als Arbeitsgrundlage haben wir zunächst eine Infrastruktur zur Datenverarbeitung geschaffen. Als Datenquellen wurden u.a. das GHTorrent-Projekt der TU Delft mit 125,5 Millionen Softwareprojekten in 1,4 Milliarden Versionen, das TravisTorrent-Projekt mit 3,7 Millionen Build-Datensätzen sowie der Stackexchange-Datensatz mit 48 Mio. Fragen und Antworten bzw. Codebeispielen herangezogen, weitere Datensätze sollen folgen. Zur Speicherung kann zwar auf existierende Datenbanksysteme zurückgegriffen werden, die nötige Softwareinfrastruktur zum Importieren, kontinuierlichen Aktualisieren, Verknüpfen, Bereinigen, und Zugänglichmachen in der Lehre der hochspezialisierten Datensätze ist hingegen von uns entwickelt worden.

Softwareentwicklung profitiert vom schnellen Auffinden von Fehlern, bevor diese später im Entwicklungs- oder Wartungsprozess teurer zu reparieren werden. Automatisierte Softwaretests helfen bei der Fehlersuche, benötigen aber häufig lange zur Ausführung. Eine KI-basierte Vorhersage kann bereits während der Änderung an der Software abschätzen, welche automatisierten Tests wahrscheinlich betroffen sind und diese zuerst ausführen, was Testzeiten im Fehlerfall deutlich verkürzt. Im Rahmen dieser Problemstellung entwickeln wir ein Verfahren zum Synthetisieren von Trainingsdaten aus der Entwicklungshistorie von Programmen und messen die Vorhersagekraft lexikalischer Programmelemente zur Fehlervorhersage. Unser Verfahren ist, im Gegensatz zu gängigen Verfahren der Testpriorisierung, auf dynamisch getypten Sprachen anwendbar. Der Anwendungsfall der Testpriorisierung ist eines der Lehrbeispiele in unserem Master-Seminar „Code Repository Mining“.

Schnittstellenentwicklung zur vereinfachten Nutzung von KI in Entwicklungswerkzeugen: Entwickler von Programmierwerkzeugen können heutige KI-Verfahren auf Quelltext häufig nicht nutzen, da eine einheitliche Schnittstelle fehlt, die Quelltext-Bestandteile mit deren Repräsentation in einem KI-Modell verbindet. In [3] untersuchen und evaluieren wir, wie eine solche Schnittstelle zu gestalten ist, damit Werkzeugentwicklung neben der formellen Struktur des Programms auch die für KI relevanten natürlichsprachlichen Elemente (z.B. Kommentare) programmatisch nutzen kann. Dies ermöglicht eine einfachere Erweiterung bestehender Entwicklungsumgebungen um KI-gestützte Codevervollständigung, Suche, Fehleranalyse, oder Tests. Ausgewählte Ergebnisse:

- Mattis, T., Rein, P., & Hirschfeld, R. (2020). Three trillion lines: infrastructure for mining GitHub in the classroom. In Proceedings of the 4th International Conference

on Art, Science, and Engineering of Programming, pp. 1-6. ACM. DOI: 10.1145/3397537.3397551.

Abstract: The increasing interest in collaborative software development on platforms like *GitHub* has led to the availability of large amounts of data about development activities. The *GHTorrent* project has recorded a significant proportion of *GitHub*'s public event stream and hosts the currently largest public dataset of meta-data about open-source development. We describe our infrastructure that makes this data locally available to researchers and students, examples for research activities carried out on this infrastructure, and what we learned from building the system. We identify a need for domain-specific tools, especially databases, that can deal with large-scale code repositories and associated meta-data and outline open challenges to use them more effectively for research and machine learning settings.

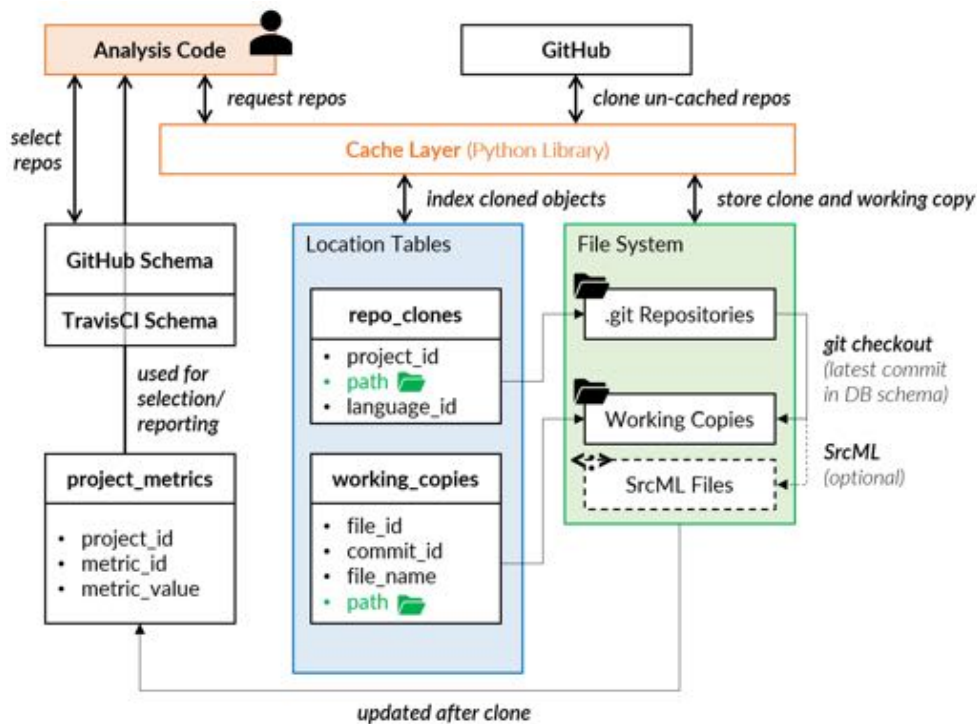


Abbildung: Systemübersicht unserer neuen Datenquelle wie sie in Forschung und Lehre eingesetzt wird.

- Mattis, T., & Hirschfeld, R. (2020). Lightweight Lexical Test Prioritization for Immediate Feedback. *The Art, Science, and Engineering of Programming*, 4(3). DOI: 10.22152/programming-journal.org/2020/4/12.

Abstract: The practice of unit testing enables programmers to obtain automated feedback on whether a currently edited program is consistent with the expectations specified in test cases. Feedback is most valuable when it happens immediately, as defects can be corrected instantly before they become harder to fix. With growing and longer running test suites, however, feedback is obtained less frequently and lags behind program changes. The objective of test prioritization is to rank tests so that defects, if present, are found as early as possible or with the least costs. While there are numerous static approaches that output a ranking of tests solely based on the current version of a program, we focus on change-based test prioritization, which recommends tests that likely fail in response to the most recent program change. The canonical approach relies on coverage data and prioritizes tests that cover the changed region, but obtaining and updating coverage data is costly. More recently, information retrieval techniques that exploit overlapping vocabulary between change and tests have proven to be powerful, yet lightweight. In this work, we demonstrate the capabilities of information retrieval for prioritizing tests in dynamic programming languages using Python as example. We discuss and measure previously understudied variation points, including how contextual information around a program change can be used, and design alternatives to the widespread TF-IDF retrieval model tailored to retrieving failing tests.

To obtain program changes with associated test failures, we designed a tool that generates a large set of faulty changes from version history along with their test results. Using this data set, we compared existing and new lexical prioritization strategies using four open-source Python projects, showing large improvements over untreated and

random test orders and results consistent with related work in statically typed languages.

We conclude that lightweight IR-based prioritization strategies are effective tools to predict failing tests in the absence of coverage data or when static analysis is intractable like in dynamic languages. This knowledge can benefit both individual programmers that rely on fast feedback, as well as operators of continuous integration infrastructure, where resources can be freed sooner by detecting defects earlier in the build cycle.

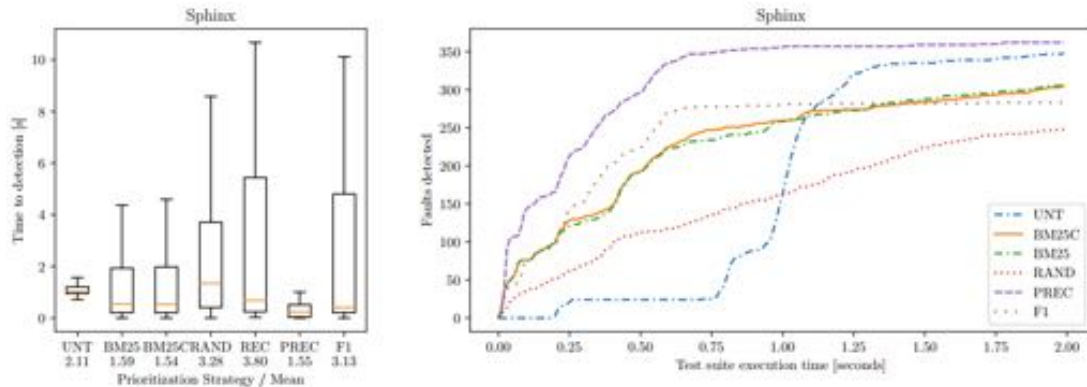


Abbildung: Mit diesem Beispiel zeigen wir die verbesserte Detektionszeit für den ersten Fehler und die kumulative Anzahl erkannter Fehler während der Testausführung für unsere neuen Verfahren (PREC) verglichen zu traditionellen (UNT/BM25) oder zufälligen (RAND) Verfahren am Beispiel eines großen Open-Source Python-Projekts.

- Mattis, T., Rein, P., & Hirschfeld, R. (2019). Ambiguous, Informal, and Unsound: Metaprogramming for Naturalness. In Proceedings of the 4th International Workshop on Meta-Programming Techniques and Reflection, pp. 1-10. ACM. DOI: 10.1145/3358502.3361270.

Abstract: Program code needs to be understood by both machines and programmers. While the goal of executing programs requires the unambiguity of a formal language, programmers use natural language within these formal constraints to explain implemented concepts to each other. This so called *naturalness* – the property of programs to resemble human communication – motivated many statistical and machine learning (ML) approaches with the goal to improve software engineering activities.

The metaprogramming facilities of most programming environments model the formal elements of a program (meta-objects). If ML is used to support engineering or analysis tasks, complex infrastructure needs to bridge the gap between meta-objects and ML models, changes are not reflected in the ML model, and the mapping from an ML output back into the program's meta-object domain is laborious.

In the scope of this work, we propose to extend metaprogramming facilities to give tool developers access to the *representations* of program elements within an exchangeable ML model. We demonstrate the usefulness of this abstraction in two case studies on test prioritization and refactoring. We conclude that aligning ML representations with the program's formal structure lowers the entry barrier to exploit statistical properties in tool development.

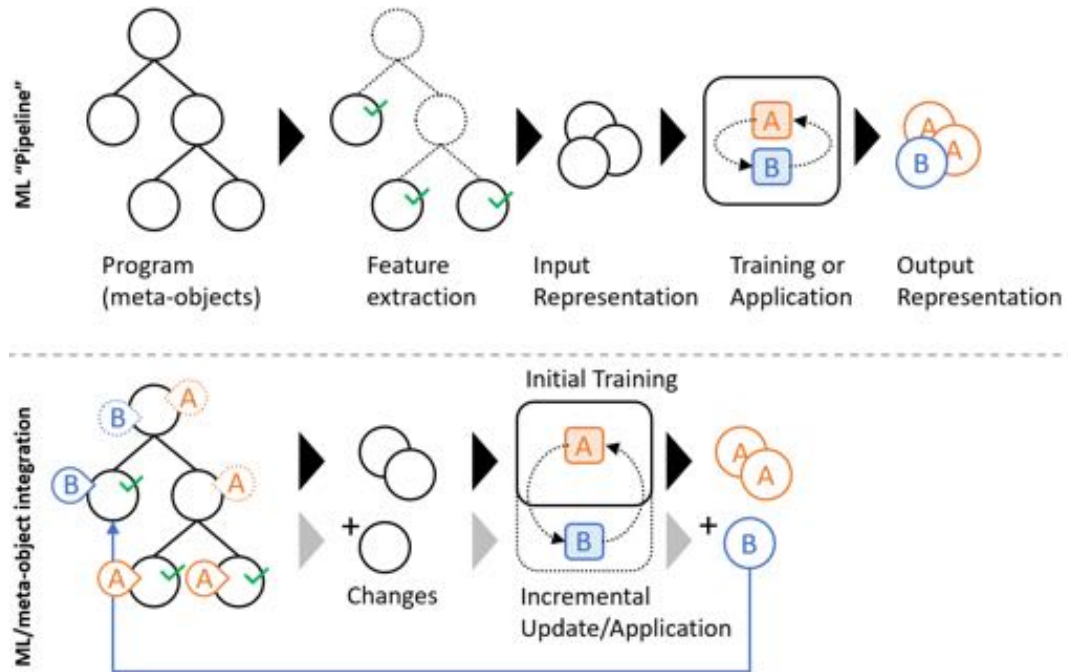


Abbildung: Inkrementelle Integration eines maschinell gelernten Modells in den Programmierfluss. Unsere Architektur (unten) berücksichtigt im Gegensatz zu klassischen "KI-Pipelines" (oben) Live-Änderungen am Programmcode.

#### 1.4.2 Arbeitspaket RH-2 "KI-basierte Entwicklungsumgebungen"

Aufbauend auf den in AP RH-1 entwickelten Verfahren wurden neue Programmierwerkzeuge für Softwareentwicklungsumgebungen (IDEs) entwickelt, deren Einsatz die Erarbeitung eines besseren Programm- und Systemverständnisses erlauben und dadurch Programmierfehlern entgegenwirken. Die besondere Herausforderung ergibt sich dadurch, dass hier eine Schnittstelle zwischen Programmierern und KI zu entwerfen ist, an die Anforderungen wie Verständlichkeit, Erklärbarkeit („Explainable AI“) und Antwortzeit/Performanz gestellt werden. Darüber hinaus muss das KI-basierte System lernfähig bleiben und skalieren können, um mit eintreffenden Änderungen vieler Programmierer gleichzeitig umzugehen (kontinuierliche Integration) sowie offline verwendbar bleiben, um den steigenden Anforderungen von Entwicklern an eine flexible Wahl ihrer Arbeitsorte gerecht zu werden.

In der planmäßigen Durchführung dieses Arbeitspakets wurden Instanzen von KI-gestützten Programmierwerkzeugen gesammelt und modelliert. Hierzu zählen insbesondere Lösungen zur Testpriorisierung, die während der Programmierfähigkeit lernen und Programmierfehler in Form fehlschlagender Tests in Echtzeit vorhersagen, sowie Code-Editoren die Programmabhängigkeiten, welche aufgrund der Projektkomplexität häufig übersehen werden, hervorheben und erklären.

Durchgeführte Lehrveranstaltungen:

- Programming Languages, Concepts, Tools, and Environments (PLCTE)
- Code Repository Mining (CRM)

- Betreuung zweier Masterarbeiten zum Thema Testpriorisierung auf historischen CI-Daten und automatisierter Modularitätseinschätzung und Messung technischer Schuld in großen Systemen

#### Ausgewählte Ergebnisse:

- Mattis, T., Rein, P., Dürsch, F., & Hirschfeld, R. (2020). RTPTorrent: An Open-source Dataset for Evaluating Regression Test Prioritization. In Proceedings of the 17th International Conference on Mining Software Repositories, pp. 385-396. IEEE/ACM. DOI: 10.1145/3379597.3387458.

Abstract: The software engineering practice of automated testing helps programmers find defects earlier during development. With growing software projects and longer-running test suites, frequency and immediacy of feedback decline, thereby making defects harder to repair. Regression test prioritization (RTP) is concerned with running relevant tests earlier to lower the costs of defect localization and to improve feedback.

Finding representative data to evaluate RTP techniques is non-trivial, as most software is published without failing tests. In this work, we systematically survey a wide range of RTP literature regarding whether their dataset uses real or synthetic defects or tests, whether they are publicly available, and whether datasets are reused. We observed that some datasets are reused, however, many projects study only few projects and these rarely resemble real-world development activity.

In light of these threats to ecological validity, we describe the construction and characteristics of a new dataset, named RTPTorrent, based on 20 open-source Java programs.

Our dataset allows researchers to evaluate prioritization heuristics based on version control meta-data, source code, and test results from fine-grained, automated builds over 9 years of development history. We provide reproducible baselines for initial comparisons and make all data publicly available.

We see this as a step towards better reproducibility, ecological validity, and long-term availability of studied software in the field of test prioritization.

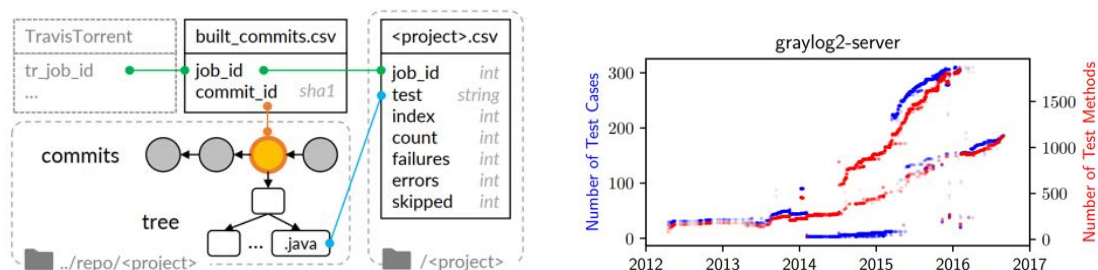


Abbildung: Architektur des RTPTorrent-Datensatzes (links) und Verlauf der Datenmenge über eines der 20 Beispielprojekt (rechts).

- Meier, D., Mattis, T., & Hirschfeld, R. (2021). Toward Exploratory Understanding of Software using Test Suites. In Proceedings of the 5th International Conference on the Art, Science, and Engineering of Programming, pp. 60-67. ACM. DOI: 10.1145/3464432.3464438.

Abstract: Changing software without correctly understanding it often leads to confusion, as developers do not understand how the change corresponds to the new observed behaviour of the system. Today, many software systems are equipped with a test suite. Test suites document code and give feedback on changed program behaviour.

We explored ways to use test suites for software comprehension and implemented a tool that provides additional visualisation and gives immediate feedback on software changes. Information about changes in the software and their implications to the test suite are collected using mutation testing. The tool uses this information to present relevant test cases for developers, and additionally prioritise test executions for immediate feedback.

Our research indicates that entropy metrics can find test cases that are relevant for a specific context in the source

code. Additionally, simple test case prioritisation strategies can already lead to a significant decrease in feedback time.

Based on our case study we argue that test suites are not only useful for regression testing but can be used to generate meaningful information for software comprehension activities.

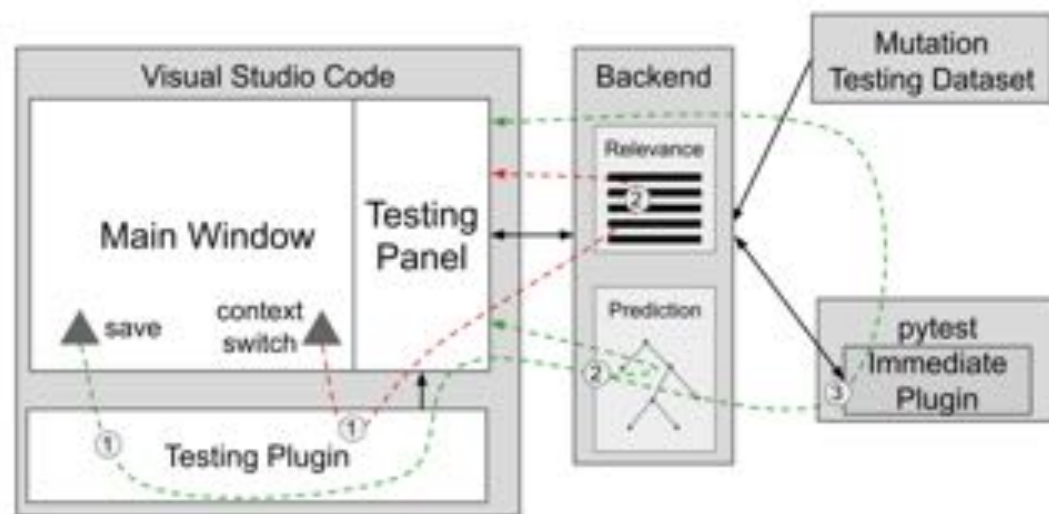


Abbildung: Architektur eines VSCode Plug-ins zum explorativen, priorisierten Testen des gerade editierten Programmcodes.

- Mattis, T., Beckmann, T., Rein, P., & Hirschfeld, R. (2021). First-class concepts: reifying architectural knowledge beyond the dominant decomposition. In Proceedings of the 13th ACM International Workshop on Context-Oriented Programming and Advanced Modularity, pp. 9-15. ACM. DOI: 10.1145/3464970.3468413.

**Abstract:** In software engineering, programs are ideally partitioned into independently maintainable and understandable modules. As a system grows, its architecture gradually loses the capability to modularly accommodate new concepts. While refactoring is expensive and the language might lack appropriate primary language constructs to express certain cross-cutting concerns, programmers are still able to explain and delineate convoluted concepts through secondary means: code comments, use of whitespace and arrangement of code, documentation, or communicating tacit knowledge.

Secondary constructs are easy to change and provide high flexibility in communicating cross-cutting concerns and other concepts among programmers. However, they have no reified representation that can be explored and maintained through tools.

In this exploratory work, we discuss novel ways to express a wide range of concepts, including cross-cutting concerns, patterns, and lifecycle artifacts independently of the dominant decomposition imposed by an existing architecture. Our concepts are first-class objects inside the programming environment that retain the capability to change as easily as code comments. We explore new tools that allow programmers to view and change programs from conceptual perspectives rather than scattering their attention across existing modules.

Our designs are geared towards facilitating multiple secondary perspectives on a system to co-exist alongside the original architecture, hence making it easier to explore, understand, and explain complex contexts and narratives not expressible in traditional modularity constructs.

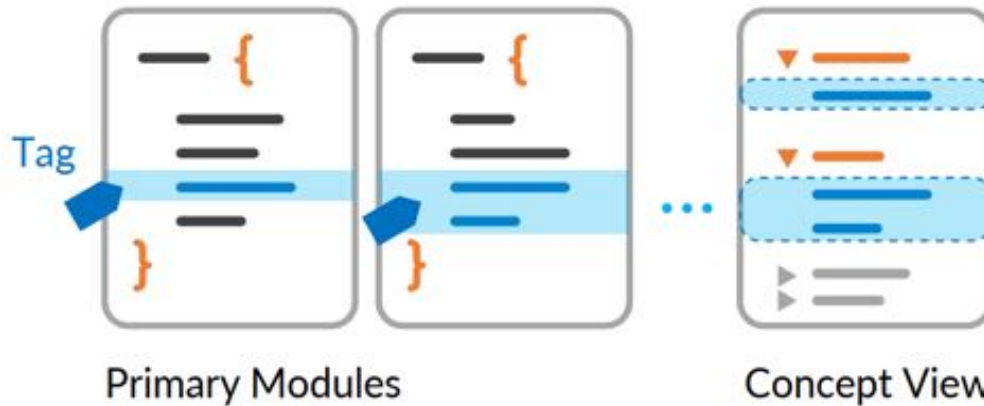


Abbildung: Design einer komponierbaren Code-Ansicht (Concept View), die maschinell gelernte querschnittende Aspekte ("Tags") aus verschiedenen Modulen übersichtlich darstellt. Programmierern wird dadurch die fehleranfällige Suche nach modulübergreifenden Zusammenhängen erleichtert und das System stellt sicher, dass Änderungen an querschnittenden Belangen in allen beteiligten Code-Modulen geschehen.

## 1.5 Arbeitsgruppe Prof. Christoph Lippert

### 1.5.1 Arbeitspaket CL1 "KI zur Modellierung von Störgrößen in hochdimensionalen Daten"

Moderne Messverfahren erlauben es, viele Millionen von molekularen Variablen auf großen Probandenkohorten zu messen. Ein entscheidender Flaschenhals in der Regressionsanalyse dieser hochdimensionalen Messungen sind unbekannte Störgrößen aufgrund von technischen Unterschieden in der Messung auf Untergruppen der Daten oder nichttechnische Struktur in der Probandenpopulation, welche statistische Abhängigkeiten in den Messungen induziert. Beispiele sind Untergruppen verschiedener genetischer Abstammung, genetische Verwandtschaften zwischen Probanden oder geteilte Umwelteinflüsse. Lineare gemischte Modelle (LMMs) sind ein etabliertes Verfahren, diese unbekanntes Störgrößen abzuschätzen und in der Regressionsanalyse zu korrigieren, haben aber limitierte Skalierbarkeit. In diesem Teilprojekt wurden Algorithmen entwickelt und evaluiert, die LMMs zur akkuraten Modellierung von unbekanntes Störgrößen in der dünn besetzten multiplen Regressionsanalyse (Lasso) auf hochdimensionalen genetischen und molekularen Daten ermöglichen. Eine Strategie ist die Entwicklung von effizienten univariaten a-priori Filter Tests, die eine Nichteinbeziehung einer gegebenen Variable in der Lösung eines LMM Lasso Regressionsmodells entweder garantieren (sichere Filter) oder enge Schranken auf die Wahrscheinlichkeit der Nichteinbeziehungen aufstellen (nichtsichere Filter).

Das Arbeitspaket wurde plangemäss durchgeführt, in dem Instanzen gesammelt und modelliert wurden. Die wissenschaftlichen Ergebnisse werden wie folgt zusammengefasst:

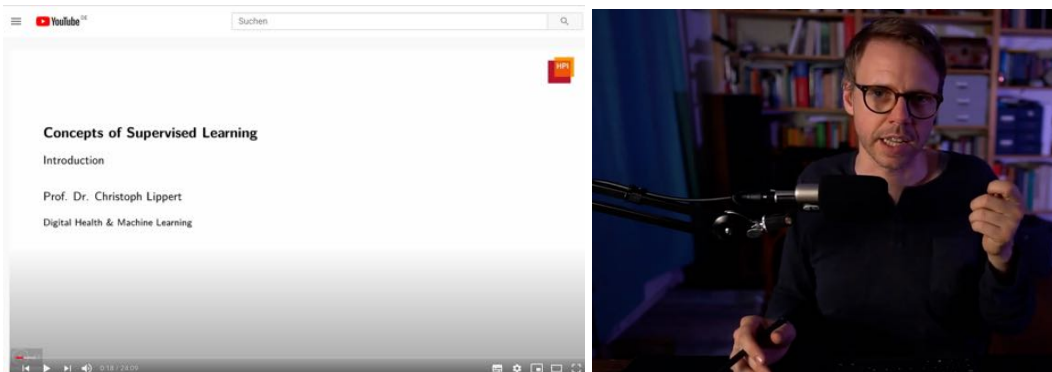
- Kirchler, M., Khorasani, S., Kloft, M., & Lippert, C. (2020). Two-sample Testing Using Deep Learning. In Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, pp. 1387-1398. PMLR.

Abstract: We propose a two-sample testing procedure based on learned deep neural network representations. To this end, we define two test statistics that perform an asymptotic location test on data samples mapped onto a hidden layer. The tests are consistent and asymptotically control the type-1 error rate. Their test statistics can be evaluated in linear time (in the sample size). Suitable data representations are obtained in a data-driven way, by solving a supervised or unsupervised transfer-learning task on an auxiliary (potentially distinct) data set. If no auxiliary data is available, we split the data into two chunks: one for learning representations and one for computing the test statistic. In experiments on audio samples, natural images and three dimensional neuroimaging data our tests yield significant decreases in type-2 error rate (up to 35 percentage points) compared to state-of-the-art two-sample tests such as kernel methods and classifier two-sample tests.

Für die massiv parallele Implementierung und Auswertung der Deep-Learning-Algorithmen wurde die spezialisierte Hardware und Speicherkapazität (ML-/DL-Hardware, GPUs und FPGAs) eingesetzt, die über die KI-Labor-Rechentechnik bereitgestellt war.

Darüber hinaus wurden themenspezifisch Vorlesungen in Deep-Learning und angewandter Mathematik (Mathematics for Machine Learning) im MOOC-Format für Bachelor- und Masterstudenten konzipiert und realisiert, die im Rahmen des KI-LAB-ITSE produziert wurden und neu entwickelte Lehrinhalte des Deep Learnings für die Medizinische Bildanalyse, Machine Learning in der Bioinformatik, sowie die statistischen Grundlagen von KI enthalten. Beispiele:

- <https://healthml.github.io/Deep-Learning-Lecture/>
- [https://www.youtube.com/playlist?list=PLkxomGYFWp65K\\_TuG2vxsnzf84WdDHLk0](https://www.youtube.com/playlist?list=PLkxomGYFWp65K_TuG2vxsnzf84WdDHLk0)
- <https://github.com/HealthML/Math4ML-Lecture>
- <https://youtube.com/playlist?list=PLkxomGYFWp67infnvPmEcqyQqk0q6ntrY>



### 1.5.2 Arbeitspaket CL2 "KI zur Genominterpretation mittels Analyse von DNS Sequenzierdaten"

Bildgebende Verfahren sind ein essenzieller Teil der klinischen Praxis. Sie unterstützen in der Vermeidung von Krankheiten, der Frühdiagnose und in der Behandlung von Krankheiten. Die Interpretierung von Bildern durch Experten in der Radiologie ist limitiert durch Subjektivität, Erschöpfung und vor allem mangelnde Skalierbarkeit der manuellen Analyse. Während in großen Gesundheitsstudien hochauflösende MRT-Bilder für zehntausende Probanden gesammelt werden, sind Expertensegmentierungen nur auf kleineren Datensätzen verfügbar. Ein Ziel dieses Projektes war es, den Prozess des

überwachten Lernens von 3D Deep-Learning-Modellen mit möglichst wenig Expertensegmentierungen zu ermöglichen. Zunächst wurden optimale Deep-Learning-Einbettungen basierend auf selbstüberwachten Lernverfahren auf großen Mengen von unannotierten MRT-Bilderdaten generiert. Diese Einbettungen wurden in weiteren Schritten mittels überwachten Transferlernen unter Benutzung der vorhandenen Annotationen verfeinert. Um die Generierung von weiteren Segmentierungen und die Korrektur von Prädiktionsfehlern zu erleichtern, wurde eine Software als semi-automatischen Annotationshilfe entwickelt.

Das Arbeitspaket wurde plangemäss durchgeführt, in dem Instanzen gesammelt und modelliert wurden. Ausgewählte Ergebnisse:

- Taleb, A., Lippert, C., Klein, T., & Nabi, M. (2019). Multimodal Self-Supervised Learning for Medical Image Analysis. arXiv preprint arXiv:1912.05396.

Abstract: In this paper, we propose a self-supervised learning approach leveraging multiple imaging modalities to increase data efficiency in machine learning for medical imaging. In particular, we introduce multimodal puzzle-solving as a proxy task to facilitate neural network feature learning from multiple image modalities with subsequent finetuning for different target tasks. To achieve that, we employ the Sinkhorn operator to predict permutations of puzzle pieces in conjunction with a modality agnostic feature embedding. Together, they allow for a lean network architecture and increased computational efficiency. Under this framework, we propose different strategies to permute medical imaging modalities, creating puzzles with varying levels of complexity. We benchmark these strategies in a range of experiments and for different target tasks. Our experiments show that solving puzzles interleaved with multimodal content yields more powerful semantic representations. This allows us to solve downstream tasks more accurately and efficiently, compared to treating each modality independently. Our approach's effectiveness is demonstrated on the semantic segmentation of brain tumor as well as survival regression task in the BraTS challenge, where we achieve results that are competitive with state-of-the-art at a fraction of the computational expense.

Für die massiv parallele Implementierung und Auswertung der Deep-Learning-Algorithmen wurde die spezialisierte Hardware und Speicherkapazität (ML-/DL-Hardware, GPUs und FPGAs) eingesetzt, die über die KI-Labor-Rechentechnik bereitgestellt war.

### **1.5.3 Arbeitspaket CL-3 "KI zur Modellierung von hochdimensionalen Krankheitsphänotypen von multimodalen Gesundheitsdaten"**

Bildgebende Verfahren und Sequenzierverfahren erlauben es, genaue Messungen von gesundheitsrelevanten Daten auf großen Kohorten zu erheben. In diesem Projekt entwickelten wir KI-Methoden, die es erlauben, statistische Zusammenhänge zwischen den Messungen aus verschiedenen nichtvektoriellen Datenmodalitäten zu entdecken. Nichtlineare Einbettungen von verschiedenen Datenmodalitäten wurden mittels Deep-Learning-Modellen gelernt. Es wurden statistische Tests entwickelt und evaluiert, die statische Abhängigkeiten in einem gemeinsamen Einbettungsraum messen.

Eine wichtige Voraussetzung für die Durchführung der Arbeiten mit ihren massiv parallelen Implementierungen und Auswertungen der Deep-Learning-Algorithmen bildet die spezialisierte projektbezogen genutzte Hardware und Speicherkapazität.

Das Arbeitspaket wurde plangemäss durchgeführt. Ausgewählte Ergebnisse:

- Taleb, A., Lippert, C., Klein, T., & Nabi, M. (2019). Multimodal self-supervised learning for medical image analysis. arXiv preprint arXiv:1912.05396. (accepted at IPMI 2021).

Abstract: Self-supervised learning approaches leverage unlabeled samples to acquire generic knowledge about different concepts, hence allowing for annotation-efficient downstream task learning. In this paper, we propose a novel self-supervised method that leverages multiple imaging modalities. We introduce the multimodal puzzle task, which facilitates rich representation learning from multiple image modalities. The learned representations allow for subsequent fine-tuning on different downstream tasks. To achieve that, we learn a modality-agnostic feature embedding by confusing image modalities at the data-level. Together with the Sinkhorn operator, with which we formulate the puzzle solving optimization as permutation matrix inference instead of classification, they allow for efficient solving of multimodal puzzles with varying levels of complexity. In addition, we also propose to utilize cross-modal generation techniques for multimodal data augmentation used for training self-supervised tasks. In other words, we exploit synthetic images for self-supervised pretraining, instead of downstream tasks directly, in order to circumvent quality issues associated with synthetic images, while improving data-efficiency and the representations learned by self-supervised methods. Our experimental results, which assess the gains in downstream performance and data-efficiency, show that solving our multimodal puzzles yields better semantic representations, compared to treating each modality independently. Our results also highlight the benefits of exploiting synthetic images for self-supervised pretraining. We showcase our approach on four downstream tasks: Brain tumor segmentation and survival days prediction using four MRI modalities, Prostate segmentation using two MRI modalities, and Liver segmentation using unregistered CT and MRI modalities. We outperform many previous solutions, and achieve results competitive to state-of-the-art.

- Kirchler, M., Konigorski, S., Norden, M. Meltendorf, C., Kloft, M. Schurmann, C., & Lippert, C. (2022). transferGWAS: GWAS of images using deep transfer learning. *Bioinformatics*, 38(14), 3621-3628. DOI: 10.1093/bioinformatics/btac369.

Abstract: Medical images can provide rich information about diseases and their biology. However, investigating their association with genetic variation requires non-standard methods. Here, we propose *transferGWAS*, a novel approach to perform genome-wide association studies with medical images as phenotypes. First, we learn semantically meaningful representations of the images based on a transfer learning task, during which a deep neural network is trained on independent but similar data. Then, we perform genetic association tests with these representations. We evaluate type I error rates and power of *transferGWAS* in simulation studies of synthetic images, and apply it to perform a genome-wide association study of retinal fundus images from the UK Biobank. This first-of-a-kind GWAS of full imaging data yielded 60 genomic regions associated with retinal fundus images, of which 7 are novel candidate loci for eye-related traits and diseases.

Darüber hinaus wurden themenspezifisch Vorlesungen zum Thema “KI in Digital Health” im Rahmen des MOOC “Digital Health für Einsteiger” auf der OpenHPI Plattform konzipiert, die im Rahmen des KI-LAB-ITSE produziert wurden und neu entwickelte Lehrinhalte zu KI zur Modellierung von hochdimensionalen Krankheitsphänotypen von multimodalen Gesundheitsdaten enthalten. (Bsp.: <https://open.hpi.de/courses/digitalhealth2020>)

#### 1.5.4 Arbeitspaket CL-4 "KI zur Genominterpretation mittels Analyse von DNS Sequenzierdaten"

Das menschliche Genom wird heute durch funktionelle Hochdurchsatz-Assays und durch die Generierung von genomischen Daten auf Populationsebene untersucht. Insbesondere Deep-Learning-Ansätze sind aufgrund ihrer Flexibilität gut geeignet, die komplexen Abhängigkeiten der regulatorischen Sequenzen des Genoms und epigenetischen Prozessen zu modellieren und Prädiktoren für die funktionale Interpretation genetischer Varianten bereitzustellen. Für die Erforschung der epigenetischen Prozesse, welche die Entwicklung von Geweben und Krebs steuern, stehen eine Vielzahl öffentlicher Datensätze zur Verfügung, die für Deep-Learning-Anwendungen im ITSE KI-Labor verfügbar gemacht

wurden. Einige der am häufigsten genutzten und größten Ressourcen sind die ENCODE und Roadmap Datensätze, welche sich auf die Entwicklung von gesunden Geweben fokussiert haben. Weiter wurden verfügbare Daten in Krebs- und Einzelzellaten aus verschiedenen Organen bearbeitet. Die Nutzbarmachung von verschiedenen Genomik-Datentypen aus verschiedenen Quellen für maschinelles Lernen beinhaltete einen beträchtlichen Aufwand. Wir haben dafür eine ressourceneffiziente Pipeline entwickelt. Wir haben eine Baseline an Modellen generiert, indem wir etablierte Modelle in der Pipeline abbildeten und verfügbar machten. Darauf aufbauend wurden eigene Modelle und Modell-Architekturen entwickelt mit dem spezifischen Fokus, pathologische Varianten im Genom zu identifizieren und zu beschreiben. Die entwickelten Modelle wurden untereinander und gegenüber publizierten Modellen verglichen und auf ihre Nützlichkeit und Skalierbarkeit für die Anwendung in der Genominterpretation evaluiert. Mit dem Ziel, die statistische Power in genomweiten Genotyp-Phänotyp Assoziationsstudien von seltenen genetischen Varianten zu erhöhen, wurden statistische Tests entwickelt, welche mittels Kernverfahren Prädiktionen zum Einfluss von genetischen Varianten auf epigenetische Prozesse als a-priori Wissen berücksichtigen. Ausgewählte Ergebnisse:

- Monti, R., Rautenstrauch, P., Ghanbari, M. L., James, A. R., Ohler, U., Konigorski, S., & Lippert, C. (2021). Identifying interpretable gene-biomarker associations with functionally informed kernel-based tests in 190,000 exomes. bioRxiv.

Abstract: Here we present an exome-wide rare genetic variant association study for 30 biomarkers in 191,640 individuals in the UK Biobank. We perform gene-based association tests for separate functional variant categories to increase interpretability and identify 201 significant gene-biomarker associations, which include novel associations such as GIGYF1 with diabetes markers. In addition to performing gene-based variant collapsing tests, we design and apply variant-category-specific kernel-based tests that integrate quantitative functional variant effect predictions for missense variants, splicing and the binding of RNA-binding proteins. For these tests we present a powerful and computationally efficient combination of the likelihood-ratio and score tests that found 32% more associations than the score test alone. Kernel-based tests identified 12-31% more associations than their gene-based collapsing counterparts with large overlaps, and had advantages in the presence of gain of function missense variants. We introduce local collapsing by amino acid position for missense variants and use this approach to identify potential novel gain of function variants in PIEZO1, and interpret a position-specific association of ABCA1-variants with inflammation marker CRP. Our results show the benefits of separately investigating different functional mechanisms when performing rare-variant association tests, and highlight the strengths of biomarker panels for large biobanks.

## 1.6 Arbeitsgruppe Prof. Meinel

### 1.6.1 Arbeitspaket CM1 "Binarisierung-KI-Konzepte"

Ein nicht gelöstes Problem von neuronalen Netzen ist der erhebliche Energie- und Rechenbedarf, der benötigt wird, um solche Modelle zu betreiben. Die Verwendung von binären Werten (statt Gleitkommazahlen) für Gewichte und Berechnungen in Modellen würde eine bis zu 32-fache Reduzierung des Energiebedarfs und der benötigten Operationen ermöglichen. In diesem Arbeitspaket wurden Methoden zum Training von binären Modellen und optimale Modellstrukturen für diese Art von Modellen untersucht. Die Genauigkeit solcher Modelle wurde anhand wissenschaftlich anerkannter Datensammlungen untersucht. Ausgewählte Ergebnisse:

- Bethge, J., Bartz, C., Yang, H., Chen, Y., & Meinel, C. (2020). MeliusNet: Can Binary Neural Networks Achieve MobileNet-level Accuracy? arXiv preprint arXiv:2001.05936.

Abstract: Binary Neural Networks (BNNs) are neural networks which use binary weights and activations instead of the typical 32-bit floating point values. They have reduced model sizes and allow for efficient inference on mobile or embedded devices with limited power and computational resources. However, the binarization of weights and activations leads to feature maps of lower quality and lower capacity and thus a drop in accuracy compared to traditional networks. Previous work has increased the number of channels or used multiple binary bases to alleviate these problems. In this paper, we instead present an architectural approach: MeliusNet. It consists of alternating a DenseBlock, which increases the feature capacity, and our proposed ImprovementBlock, which increases the feature quality. Experiments on the ImageNet dataset demonstrate the superior performance of our MeliusNet over a variety of popular binary architectures with regards to both computation savings and accuracy. Furthermore, with our method we trained BNN models, which for the first time can match the accuracy of the popular compact network MobileNet-v1 in terms of model size, number of operations and accuracy.

Für die Implementierung der Demonstratoren und Frameworks sowie für experimentelle Arbeiten und Auswertung der Ansätze wird insbesondere die spezialisierte Hardware und Speicherkapazität eingesetzt, die über die KI-Labor-Rechentechnik bereitgestellt wird.

- Bethge, J., Bartz, C., Yang, H., Chen, Y., & Meinel, C. (2020). MeliusNet: Can Binary Neural Networks Achieve MobileNet-level Accuracy? arXiv preprint arXiv:2001.05936.

Abstract: Binary Neural Networks (BNNs) are neural networks which use binary weights and activations instead of the typical 32-bit floating point values. They have reduced model sizes and allow for efficient inference on mobile or embedded devices with limited power and computational resources. However, the binarization of weights and activations leads to feature maps of lower quality and lower capacity and thus a drop in accuracy compared to traditional networks. Previous work has increased the number of channels or used multiple binary bases to alleviate these problems. In this paper, we instead present an architectural approach: MeliusNet. It consists of alternating a DenseBlock, which increases the feature capacity, and our proposed ImprovementBlock, which increases the feature quality. Experiments on the ImageNet dataset demonstrate the superior performance of our MeliusNet over a variety of popular binary architectures with regards to both computation savings and accuracy. Furthermore, with our method we trained BNN models, which for the first time can match the accuracy of the popular compact network MobileNet-v1 in terms of model size, number of operations and accuracy.

- Bethge, J., Bartz, C., Yang, H., Chen, Y., & Meinel, C. (2021). MeliusNet: An Improved Network Architecture for Binary Neural Networks. In Proceedings of the Winter Conference on Applications of Computer Vision, pp. 1439-1448. IEEE/CVF.

Abstract: Binary Neural Networks (BNNs) are neural networks which use binary weights and activations instead of the typical 32-bit floating point values. They have reduced model sizes and allow for efficient inference on mobile or embedded devices with limited power and computational resources. However, the binarization of weights and activations leads to feature maps of lower quality and lower capacity and thus a drop in accuracy compared to their 32-bit counterparts. Previous work has increased the number of channels or used multiple binary bases to alleviate these problems. In this paper, we instead present an architectural approach: MeliusNet. It consists of alternating a DenseBlock, which increases the feature capacity, and our proposed ImprovementBlock, which increases the feature quality. Experiments on the ImageNet dataset demonstrate the superior performance of our MeliusNet over a variety of popular binary architectures with regards to both computation savings and accuracy. Furthermore, BNN models trained with our method can match the accuracy of the popular compact network MobileNet-v1 in terms of model size and number of operations.

Für die Implementierung der Demonstratoren und Frameworks sowie für experimentelle Arbeiten und Auswertung der Ansätze wurde insbesondere die spezialisierte Hardware und Speicherkapazität eingesetzt, die über die KI-Labor-Rechentechnik bereitgestellt wird.

### 1.6.2 Arbeitspaket CM-2 "Weakly-supervised Training des neuronalen Netzwerks"

Heutige Deep Learning basierte Verfahren, z.B. zur Lokalisierung und Erkennung von Texten oder Objekten in natürlichen Szenenbildern, beruhen auf der Verfügbarkeit von Millionen von vollständig annotierten Datenpunkten. Die Erstellung eines solchen Datensatzes mit Millionen von Annotationen ist nicht nur sehr zeitaufwändig, sondern benötigt auch beträchtliche finanzielle Ressourcen. Dieses Arbeitspaket konzentrierte sich auf die Entwicklung neuer Ansätze, die auf schwach überwachten Methoden basieren. Mithilfe dieser schwachen Überwachung ist es möglich, Systeme zu trainieren, die einzelne Teilaufgaben, für die heutzutage noch Annotationen benötigt werden, ohne weitere Annotation der Daten zu lösen. Dazu wurden geeignete Verfahren entwickelt und ausgiebig getestet. Ausgewählte Ergebnisse:

- Bethge, J., Bartz, C., Yang, H., Chen, Y., & Meinel, C. (2021). MeliusNet: An Improved Network Architecture for Binary Neural Networks. In Proceedings of the Winter Conference on Applications of Computer Vision, pp. 1439-1448. IEEE/CVF.

Abstract: Binary Neural Networks (BNNs) are neural networks which use binary weights and activations instead of the typical 32-bit floating point values. They have reduced model sizes and allow for efficient inference on mobile or embedded devices with limited power and computational resources. However, the binarization of weights and activations leads to feature maps of lower quality and lower capacity and thus a drop in accuracy compared to their 32-bit counterparts. Previous work has increased the number of channels or used multiple binary bases to alleviate these problems. In this paper, we instead present an architectural approach: MeliusNet. It consists of alternating a DenseBlock, which increases the feature capacity, and our proposed ImprovementBlock, which increases the feature quality. Experiments on the ImageNet dataset demonstrate the superior performance of our MeliusNet over a variety of popular binary architectures with regards to both computation savings and accuracy. Furthermore, BNN models trained with our method can match the accuracy of the popular compact network MobileNet-v1 in terms of model size and number of operations.

- Guo, N., Bethge, J., Yang, H., Zhong, K., Ning, X., Meinel, C., & Wang, Y. (2021). Boolnet: minimizing the energy consumption of binary neural networks. arXiv preprint arXiv:2106.06991.

BoolNet ist eine Arbeit, die sich auf das Gleichgewicht zwischen Stromverbrauch und Genauigkeit konzentriert. BoolNet optimiert und entfernt fast alle 32-Bit-Recheneinheiten basierend auf dem Designkonzept von MeliusNet weiter und begrenzt den Informationsfluss innerhalb des Netzwerks auf 1-Bit. Da ein größerer Anteil des Engpasses beim Stromverbrauch von KI-Beschleunigern eher vom Speicherzugriff als vom parallelen Rechnen herrührt. Daher reduziert der von BoolNet entworfene 1-Bit Informationsfluss den Stromverbrauch erheblich, indem der Speicherzugriff reduziert wird. Gleichzeitig reduziert BoolNet durch den Einsatz fortschrittlicher Optimierungstechniken um das 4,6-fache weniger Stromverbrauch im Vergleich zum state-of-the-art BNN-Modell bei minimalem Genauigkeitsverlust. Diese Arbeit erfolgt durch gemeinsames Design von Algorithmen und Hardware. Die Stromverbrauchsmessungen wurden durch die ASIC-basierte KI-Beschleuniger-Simulationsverifizierung von unseren Partner aus Nanoscale Integrated Circuits and Systems Lab der Tsinghua University erhalten.

- Bartz, C., Seidel, L., Nguyen, D.-H., Bethge, J., Yang, H., & Meinel, C. (2020). Synthetic Data for the Analysis of Archival Documents: Handwriting Determination. In

Proceedings of the Digital Image Computing: Techniques and Applications, pp. 1-8. IEEE. DOI: 10.1109/DICTA51227.2020.9363410.

Abstract: Archives contain a wealth of information and are invaluable for historical research. Thanks to digitization, many archives are preserved in a digital format, making it easier to share and access documents from an archive. Handwriting and handwritten notes are commonly found in archives and contain a lot of information that can not be extracted by analyzing documents with Optical Character Recognition (OCR) for printed text. In this paper, we present an approach for determining whether a scan of a document contains handwriting. As a preprocessing step, this approach can help to identify documents that need further analysis with a full recognition pipeline. Our method consists of a deep neural network that classifies whether a document contains handwriting. Our method is designed in such a way that we overcome the most significant challenge when working with archival data, which is the scarcity of annotated training data. To overcome this problem, we introduce a data generation method to successfully train our proposed deep neural network. Our experiments show that our model, trained on synthetic data, can achieve promising results on a real-world dataset from an art-historical archive.

- Hu, T., & Meinel, C. (2020). Text Generation in Discrete Space. In Proceedings of the International Conference on Artificial Neural Networks, pp. 721-732. Springer.

Abstract: Variational AutoEncoders (VAEs) are applied to many generation tasks while suffering from posterior collapse issue. Vector Quantization (VQ) is recently employed in VAE model on image generation, which could get rid of the posterior collapse problem and show its potentiality for more generation tasks. In this paper, the VQ method is applied to VAE on text generation. We elaborately design the model architecture to mitigate the index collapse issue brought in by VQ process. Experiments show that our text generation model can achieve better reconstruction and generation performance than other VAE based approaches.

- Eslami, S., Adorján, P., & Meinel, C. (2020). SehMIC: Semi-hierarchical Multi-label ICD Code Classification. In Proceedings of the CEUR Workshop.

Für die Arbeiten wurde insbesondere spezialisierte Hardware (ML-/DL-Hardware, GPU-Cluster) eingesetzt; für die Speicherung der Bilddaten für Training und Testing wurde Speicherkapazität (Storage Server) genutzt. GPU-Computing-Services, sowie FPGAs und Kleingeräte für KI-Anwendungen wurden experimentell für die Ausführung der Verfahren in verschiedenen Kontexten eingesetzt, um die Skalierbarkeit für den Einsatz im Rechenzentrum, sowie die Einsatzmöglichkeiten direkt auf Edge Devices (z.B. Mobilgeräten, oder IoT Geräten) zu ermitteln.

## 1.7 Übergreifende Maßnahmen/KI-Labor Allgemein

### 1.7.1 Arbeitspaket L1 "KI-Labor-Hardware"

Dieses Arbeitspaket befasste sich generell mit Laboraufgaben im Bereich der Hardware und Software, die spezifisch für das KI-Labor eingesetzt wurden. Die Arbeiten umfassten die Inbetriebnahme, Konfiguration und fortlaufende Kontrolle der dezidierten KI-Labor-Hardware und die Inbetriebnahme, Lizenzierung und Konfiguration KI-spezifischer Datensammlungen, Softwarebibliotheken und Software-Frameworks. Tätigkeiten und Ergebnisse:

- Analyse der konkreten Anforderung an die Rechentechnik innerhalb der KI-Labor-Forschung am HPI
- Aufbereitung der Anforderungen zu technischen Spezifikationen für die KI-Labor-Rechentechnik am HPI

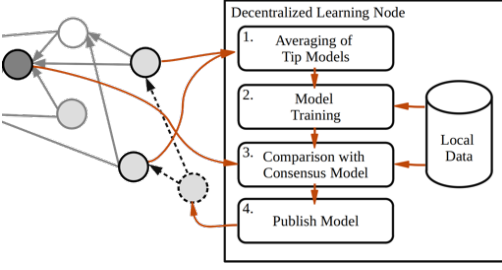
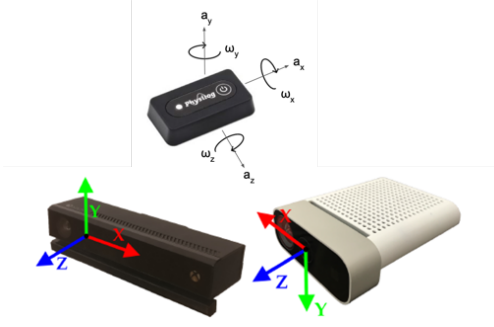
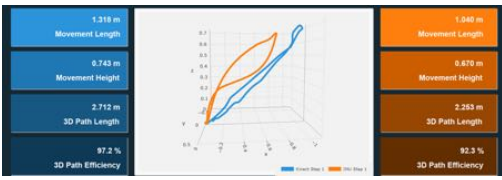
- Angebotsermittlung und Rücksprache mit Lösungsanbietern; Abstimmung mit Projektträger
- Beratung und Konzepterstellung für die Integration der KI-Labor-Rechentechnik in die HPI-eigene Rechenzentrums Umgebung
- Fachkonzept für die integrierte Rechentechnik im (im Aufbau befindlichen) HPI-Rechenzentrum
- Fachliche Prüfung und Diskussion der technischen Umsetzung der Hardware-Anforderung; Beauftragung zu "Storage" und "Server"
- Einrichtung und (fortlaufende) Konfiguration der Hardware-Komponenten
- Einrichtung und (fortlaufende) Konfiguration der Software-Komponenten und Software-Frameworks sowie der erforderlichen Testdatenbibliotheken
- Anforderungsanalyse, Planung und Angebotserstellung zu fachspezifischen Hardware- und Softwarebedarf der KI-Labor-Teilbereiche
- Anforderungsanalyse, Planung, Angebotserstellung, Auswertung und Beauftragung zu Trainingsdaten für KI-Verfahren; Einrichtung und Übernahme der Testdaten
- Beauftragung, Einrichtung und Konfiguration fachspezifischer Komponenten und Dienste, insbesondere Softwareentwicklungsplattformen und KI-Services
- Einrichtung von spezifischer Labor-Hardware, z.B. für Labortisch und Bauteile für Sensor-Netzwerke "KNX"



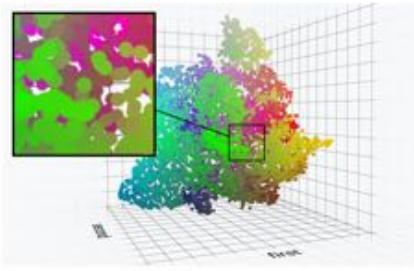
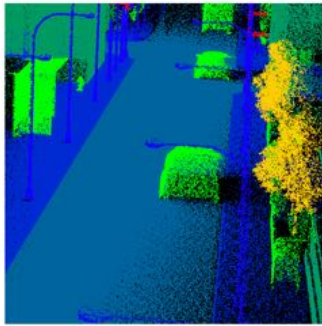
#### **Meilenstein M1.1:**

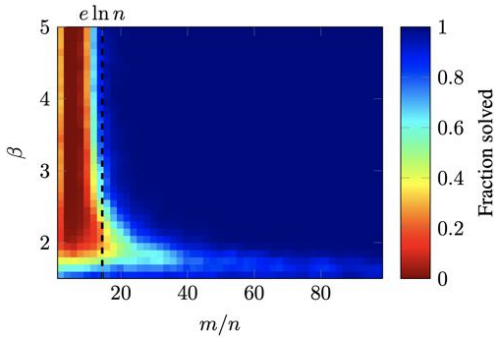
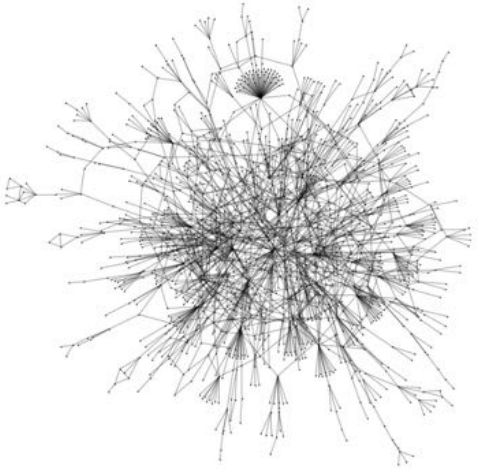
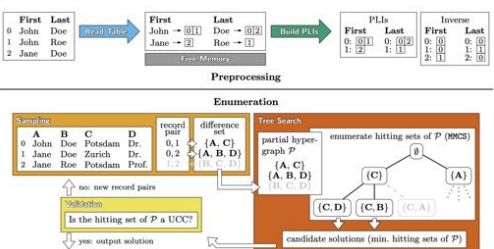
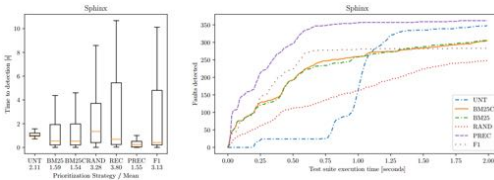
- Inbetriebnahme KI-Labor-Hardware/-Software erfolgte innerhalb des HPI-Rechenzentrums sowie in den einzelnen fachlich zuständigen Fachgebieten am HPI
- Einrichtung, fortlaufende Rekonfiguration und Optimierung der Hardware-Komponenten im KI-Labor
- Fortlaufende technische Durchsicht der laufenden Beschaffungen auf technische Kompatibilität mit der bisherigen KI-Labor-Rechentechnik
- Betriebsunterstützung von spezifischer Labor-Hardware, z.B. für Labortisch und Bauteile für Sensor-Netzwerke "KNX"
- Einrichtung, fortlaufende Rekonfiguration und Optimierung der Speicher-Komponenten, einschließlich der mobilen Speicher für den Austausch größter Datensätze (Big Data) für Aufgaben des KI-Labors (z.B. KNX-Datensätze, 3D-Punktwolken-Datensätze)
- Analyse der konkreten Anforderung bzgl. Ergänzungs- und Sonderbeschaffungen
- Angebotsermittlung, Rücksprache mit Lösungsanbietern für Software-Frameworks; Abstimmung mit Projektträger. Einrichtung und Betrieb der Software-Frameworks
- Angebotsermittlung, Rücksprache mit Lösungsanbietern für KI-relevante Test- und Trainingsdaten; Abstimmung mit Projektträger. Einrichtung und Bereitstellung der Daten
- Beauftragung, Einrichtung und Konfiguration fachspezifischer Komponenten und Dienste, insbesondere Softwareentwicklungsplattformen und KI-Services
- Einrichtung und fortlaufende Konfiguration der Software-Komponenten und Software-Frameworks (z.B. KI-Framework Digital Masterpieces) sowie der erforderlichen Testdatenbibliotheken (z.B. KNX-Testdatenstellung)

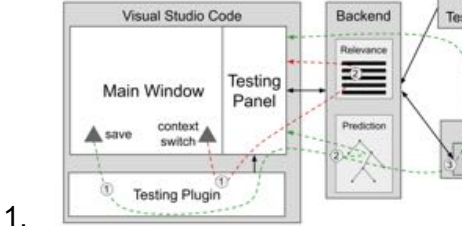
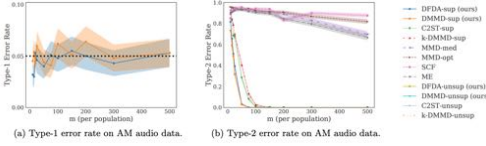
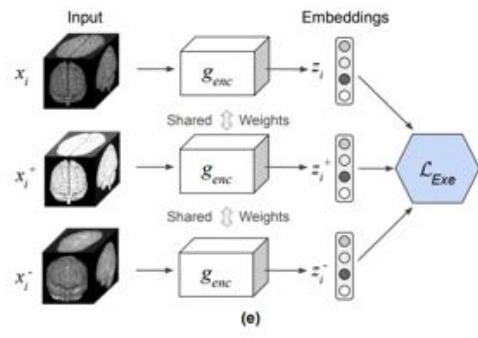
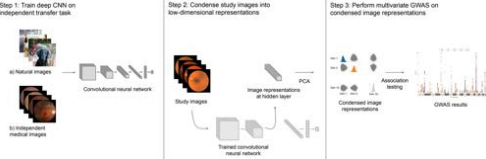
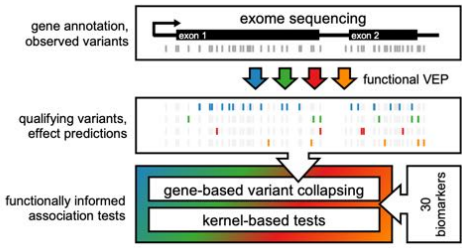
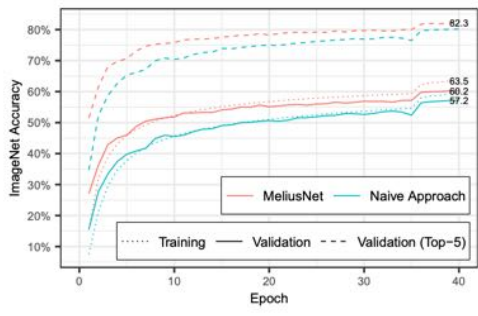
- Auswahl und Konfiguration von Werkzeugen und Diensten für statische Informationsvisualisierung, z.B. zur Qualitätskontrolle von Trainingsdaten, Testdaten und Ergebnisdaten, u.a. auf Basis von Tree Maps und Topic Maps
- Auswahl und Konfiguration von Verfahren für dynamische Informationsvisualisierung, z.B. zur Qualitätskontrolle von Trainingsdaten, Testdaten und Ergebnisdaten, u.a. auf Basis von Partikel-basierten Visualisierungen wie z.B. Dust-&-Magnet
- Recherche und Verwaltung von anwendungsgebietspezifischen Cloud-Services, z.B. für Geo- oder Bild-Trainingsdaten und -Trainingsnetze

**Meilenstein M1.2: Snapshots von Demonstratoren**

AP	Thema	Demonstrator	Kommentar
BA-1/2 /3/4	Federated Learning		<p>Der dezentrale Ansatz für das verteilte Lernen mithilfe eines gerichteten azyklischen Graphen ermöglicht ein schnelleres und sichereres Training von maschinellen Lernmodellen.</p>
	Internet-of-Health-Things		<p>Zwei Methoden zur Erfassung von Bewegungsdaten wurden evaluiert: Computer Vision (CV) und Beschleunigungssensoren (IMU). Wir verglichen die Datenqualität der in die engere Wahl gekommenen kommerziellen Messgeräte, legten die Geräte für die Datenerfassung fest und setzten Methoden des maschinellen Lernens ein, um unsere Daten im Kontext der Ganganalyse, des Krafttrainings und der Erkennung menschlicher Aktivitäten auszuwerten.</p>
	Connected Health Patientenakte		<p><i>Demonstrator: Auswirkungen des stationären Fastens auf Blutparameter.</i></p>
	Transparente Ki		<p>Dashboard-App für KI-basierte Datenanalyse (Fussbewegung)</p>

<p>JD-1/2/ 3/4</p>	<p>KI-basierte Neural Style Transfer Verfahren</p>		<p>Forschungsergebnisse des HPI im Bereich Neural Style Transfer wurden von der Graphite App aufgegriffen und sind dort "live" erlebbar. Die prototypische Implementierung erfolgt auf Basis der KI-Softwareentwicklungs-Plattform von Digital Masterpieces.</p>
	<p>KI-basierte Per-Frame Video Analyse</p>		<p>Forschungsprototyp einer KI-basierten Video-Analyse. Im Beispiel werden die von Personen ausgeführten Bewegungsabläufe im Video identifiziert und dann mit grafischen Verfahren symbolisiert. Dazu werden die abgeleiteten Bewegungsinformation zeitlich kohärent berechnet.</p>
	<p>KI für Softwareanalytics</p>		<p>3D-Scatter-Plot Demonstrator-Software</p>
	<p>KI für 3D-Punktwolken</p>	 <p>(b) Close up view.</p>	<p>3D-Punktwolken-Simulator</p>

<p>TF-12/ 3/4</p>	<p>Testdatensätze für SAT</p>		<p>Es wurden Testdatensätze für SAT gesammelt. Dargestellt ist die Untersuchung der Erfüllbarkeit von skalenfreien 3-SAT Formeln. Visualisiert ist der Anteil der Formeln, die der Greedy Algorithmus von Koutsoupias-Papadimitriou (1992) löst, abhängig von der Constraint Dichte <math>m/n</math> und dem Power-Law-Exponent <math>\beta</math>.</p>
	<p>Testdatensätze für Netzwerke</p>		<p>Es wurden Netzwerke aus ganz verschiedenen Domänen zusammengetragen. Dargestellt ist eine hyperbolische Einbettungen der größten Komponente (1458 Knoten) des Netzwerkes "bio-yeast-protein-inter" mit dem Spring Embedder von Hachul-Jünger.</p>
	<p>HPIValid</p>		<p><i>Aufbau unseres neuen Algorithmus HPIValid. Im Preprocessing werden die Tabellen gelesen und eine vorläufige Clusterstruktur extrahiert. Über eine eigene Datenstruktur wird anschließend gesampelt und eine Baumsuche durchgeführt, um gültige Lösungen zu sammeln und validieren.</i></p>
<p>RH-1/2</p>	<p>Predictive Software Maintenance</p>		<p>Mit diesem Beispiel zeigen wir die verbesserte Detektionszeit für den ersten Fehler und die kumulative Anzahl erkannter Fehler während der Testausführung für unsere neuen Verfahren (PREC), verglichen zu traditionellen (UNT/BM25) oder zufälligen (RAND) Verfahren am Beispiel eines großen Open-Source Python-Projekts.</p>

	<p>KI-basierte Entwicklungsumgebungen</p>	 <p>1.</p>	<p>VSCode-Plugin Demonstrator</p>
<p>CL-1,2, 3,4</p>	<p>KI zur Modellierung von Störgrößen</p>	 <p>(a) Type-1 error rate on AM audio data. (b) Type-2 error rate on AM audio data.</p>	<p>Demonstrator zur Modellierung von Störgrößen (hier: Audio-Daten).</p>
	<p>KI zur Interpretation, Segmentierung und Klassifikation von hochauflösenden 3D medizinischen bildgebenden Verfahren</p>	 <p>(e)</p>	<p>Demonstrator zur Interpretation von medizinischen 3D-Bildern.</p>
	<p>KI-Modellierung hochdim. Krankheits-Phänotypen</p>		<p>transferGWAS Demonstrator</p>
	<p>KI zur Genominterpretation</p>		<p>Demonstrator "Rare-Variante Association Testing Pipeline"</p>
<p>CM-1/2</p>	<p>KI-Konzepte für Binarisierung</p>		<p>The MeliusNet Demonstrator for Binary Neural Networks.</p>
		<p>Bitorch: An Open Source Binary Neural Network Framework Link: <a href="https://github.com/hpi-xnor/bitorch">https://github.com/hpi-xnor/bitorch</a></p>	<p>OpenSource-Framework Entwicklung zum Them Neural Networks</p>

		<p>Abstract: BITorch ist eine Bibliothek, die derzeit entwickelt wird, um den Aufbau quantisierter und binärer neuronaler Netze mit PyTorch zu vereinfachen. Die aktuelle Roadmap von Bitorch enthält: Erweiterung des Modellzoos mit vortrainierten Modellen nach State-of-the-Art-Ansätzen Hinzufügen von Beispielen für fortgeschrittene Trainingsmethoden mit mehreren Stufen, Wissensdestillation usw.; Integration der Bitorch-Engine für beschleunigte Inferenz auf GPU und CPU.</p>	
--	--	---	--

### 1.7.2 Arbeitspaket L2 "KI-Labor-MOOCs"

Dieses Arbeitspaket befasste sich mit der Vorbereitung, Koordinierung und Erfassung der KI-bezogenen Aus- und Weiterbildungsinhalte, die aus den fachlichen Arbeitspaketen der beteiligten Fachgebiete entstehen, und mit der Vorbereitung und Recherche für den Know-How-Transfer sowie mit dem Aufbau zu Wirtschafts- und Verwaltungskontakten bzgl. des KI-Labors. Zum einen umfasste dies Massive Open Online Courses (MOOCs) und Teletask-Kurse, die sich wegen der pandemiebedingten digitalen Umstellung der Lehre als besonders wichtiges Format für die akademische Lehre entwickelt haben. Zum anderen beinhaltete es qualifizierte Aus- und Weiterbildungsformate mit verschiedenen Akteuren in der Wirtschaft.

#### MOOC-Konzepte und Technologien für das KI-Labor

- Taheri, M., Hölzle, K., & Meinel, C. (2019). Towards Culturally Inclusive MOOCs: A Design-based Approach. CSEDU (1), 597-604.

Abstract: Massive Open Online Courses (MOOCs) have become one of the most popular ways of acquiring new knowledge and skills. One of the unique characteristics of MOOCs is their learner diversity. MOOC learners vary in age, gender, cultural background, language and discipline. This poses a great challenge for MOOC designers to create learning experiences that resonate with their diverse global audience. This paper reports instructional strategies that were applied to create culturally inclusive MOOCs. We applied a design-based approach to experiment, test, and evaluate these strategies over the course of three MOOCs on the topic of Design Thinking. The study uses in-depth qualitative interviews with international participants, pre and postcourse surveys, as well as observations from the discussion forums, in order to gain insights into learners' perspectives. As a result, the authors offer instructional strategies that may accommodate the needs of MOOC learners from diverse cultural backgrounds. Considering that MOOCs promise opportunities for life-long learning to learners around the globe, it is of utmost importance to design learning experiences that are culturally inclusive.

- Serth, S., Staubitz, T., Teusner, R., Meinel, C (2021). CodeOcean and CodeHarbor: Auto-Grader and Code Repository. In Proceedings of the 52nd Technical Symposium on Computer Science Education. ACM.

Abstract: The Hasso Plattner Institute (HPI) successfully operates a MOOC (Massive Open Online Course) platform since 2012. Since 2013, global enterprises, international organizations, governments, and research projects funded by the German ministry of education are partnering with us to operate their own instances of the platform. The focus of our platform instance is on IT topics, which includes programming courses in different programming languages. An important element of these courses are graded hands-on programming assignments. MOOCs, even more than traditional classroom situations, depend on automated solutions to assess programming exercises. Manual evaluation is not an option due to the massive amount of users that participate in these courses. The paper at hand presents two of the tools developed in this context at the HPI: CodeOceantextemdash an auto-grader for a variety of programming languages, and CodeHarbor, a tool to share auto-gradable programming exercises between various online platforms.

- Bothe, M., & Meinel, C. (2020). When Do Learners Rewatch Videos in MOOCs?. In Proceedings of Learning with MOOCs, pp. 148-151. IEEE. DOI: 10.1109/LWMOOCS50143.2020.9234368.

Abstract: Mobile applications for MOOCs (Massive Open Online Courses) offer the possibility to download learning material to enable network independent learning sessions. The management of downloaded content on mobile devices is a manual process for the learner, which has the potential for automation. This includes the deletion of learning material that is likely to be no longer consumed. In this paper, a metric was defined to quantify learners' references to previous videos based on the order in which the videos were viewed. In an observational study involving three MOOCs in the field of computer science and IT systems engineering, learners referred to previous video content only a single time on average. Outliers made use of earlier content up to 44 times during a course. Referenced videos belonged in most cases to the current or previous course section. The learners referred more often to previous videos during the course period compared to when participating in self-paced mode, while learners who earned a record of achievement referred to previous videos significantly more frequently than those who did not.

- Meinel, C. u.a.: Applied Edge AI: Deep Learning outside of the Cloud, openHPI, 2022, <https://learn.ki-campus.org/courses/edgeai-hpi2022/>, <https://open.hpi.de/courses/edgeai2022>

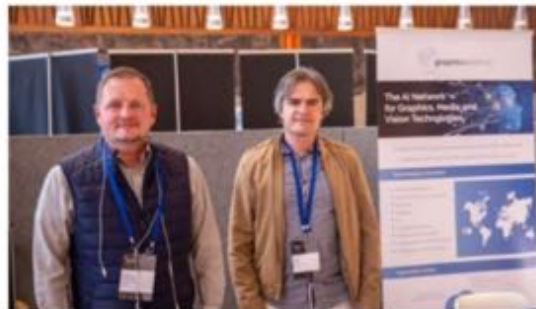
Abstract: Die inzwischen immer bekannter werdenden Techniken des Deep Learning werden inzwischen von immer mehr Unternehmen und Forschern für verschiedene Anwendungen in den Bereichen der Technik, Medizin, und Online-Kommerz verwendet. Häufig wird das Modell dabei in der Cloud ausgeführt. Das erfordert Übertragung von Daten, oft werden dabei zentrale Datensammlungen angelegt und die Verarbeitung ist von einer stabilen Anbindung an die Cloudressourcen abhängig. Eine Alternative dazu besteht in Ansätzen der "Edge AI". Hier werden Daten dezentral und nah an der Quelle verarbeitet. Dadurch können unnötige Kopien, Bedenken gegen Privatsphäre, und zwingende Cloud Anbindung vermieden werden. Edge AI verarbeitet dabei Daten direkt auf Endgeräten, zum Beispiel, Smartphones, Messgeräten in Fabriken, oder medizinischen Geräten und kann so zu höherer Effizienz, niedrigen Latenzen und besserem Datenschutz führen. Der englischsprachige Kurs wiederholt dabei für Interessierte, die bereits erste Erfahrungen mit Deep Learning und der Programmiersprache Python gemacht haben, zunächst die Grundlagen von Deep Learning, um dann tiefer in die Details zu verschiedenen Ansätzen zu gehen. Eine besondere Herausforderung und wertvolle Lernerfahrung sind dabei die praktischen Programmierübungen in Python mit den Bibliotheken Numpy und PyTorch, die im Laufe des Kurses bearbeitet werden. Am Ende des Kurses werden die Teilnehmenden einen guten Überblick über die verschiedenen Ansätze von Edge AI haben und nicht nur die Theorie verstanden haben, sondern auch praktische Programmiererfahrung geübt haben. Für die Teilnahme am Kurs ist ein wöchentlicher Zeitaufwand von 6-8 Stunden einzuplanen.

### **Aus- und Weiterbildung für Partner aus Wirtschaft und Verwaltung:**

- Koordination und Kommunikation mit dem Verband der Geoinformationswirtschaft GeoKomm. Im Mittelpunkt stand die erfolgreiche Gründung des ZIM-Fördernetzwerks „risKI – Resilienz, Integrität und Sicherheit kommunaler Infrastrukturen“ (2022) zum Thema Künstliche Intelligenz für räumliche Umgebungen. Es hat bisher 12 Partner aus Wirtschaft, Verwaltung und Hochschule. Das Netzwerk ermöglicht es sehr gut, das KI-Knowhow aus dem KI-Labor für diese spezifischen Fragestellungen der Wirtschaft und Verwaltung zugänglich zu machen. Es hat den Schwerpunkt auf KI-Anwendungen und KI-Services, die räumliche, intelligente digitale Zwillinge für Umgebungen (Indoor und Outdoor) bereitstellen.
- Koordination und Kommunikation mit dem Fachverband GraphicsVision.AI: Eine Konzeptentwicklung für einen Research Retreat zum Thema "AI & Visual Computing" wurde abgeschlossen und soll innerhalb von GraphicsVision.AI ab Q3/2022

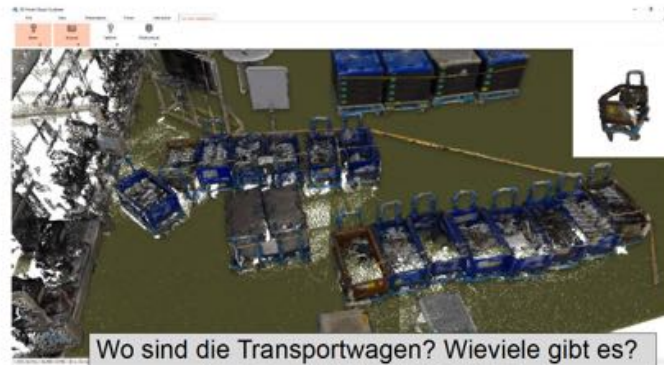
anlaufen. Dieses Format dient zum Know-How-Transfer aus dem KI-Labor auf KMUs innerhalb der KMU-Mitglieder des GraphicsVision.AI-Netzwerks sowie allgemein der KI-affinen Wirtschaft. Das HPI hat dazumit GraphicsVision.AI die Vorbereitungen für einen ersten KI-Labor-Planungsworkshop vorgenommen (Konf. ICSSA, Málaga, Juli 2022; Institute for Domotics and Energy Efficiency, Univ. Málaga).

- Im KI-Labor wurde mit der Universidad de Monterrey, México (“Tecnológico de Monterrey”, TEC; Prof. Valle) ein Memorandum of Understanding erarbeitet, das KI als zentrales Thema beinhaltet. Derzeit wird ein Konzept zwischen HPI und TEC entwickelt, wie Lehre und Forschung zwischen den Partnern zukünftig gestaltet und praktisch durchgeführt werden können. Im Mittelpunkt stehen anfänglich MOOCs zum Thema KI und die gezielte Anwerbung von TEC-Studierenden, die sich auf KI am HPI spezialisieren wollen (Masterarbeiten, Dissertationen).
- Präsentation des KI-Labors auf dem EU Robotics Forum im März 2020 in Malaga, Spanien durch Prof. Döllner und den Verband GraphicsVision.AI. Co-Hosting von Workshops, Messestand mit Präsentation des KI-Labors (s. Foto). Organisation und Präsentationen durch Prof. Döllner, J.F. Carvajal und J. Bund. Im Nachgang konnte GraphicsVision.AI systematisch potentielle Anwender ausgewählter KI-Technologie ermitteln, ansprechen und erste Kontakte für etwaige spätere gemeinsame Fortführungsprojekte herstellen.



- Vorbereitung eines Workshops für AI und Geospatial Data mit der Universidad de Sevilla, KI-Labor und GraphicsVision.AI Q4/2022.
- Ausarbeitung einer Studie zur AI-basierten Analyse von 3D-Punktwolken als Basisdaten für räumliche digitale Zwillinge gemeinsame mit der Volkswagen AG Foto: AI eingesetzt zur Interpretation nicht strukturierter, nicht semantisch attributierter 3D-Punktwolken zur Objektidentifikation und -separation (im Beispiel Erkennen von Transportwagen in der 3D-Punktwolken).

## Objekterkennung (Beispiel 1)



**Meilenstein M2.1:** (3. Monat nach Projektbeginn): Index und Themenübersicht KI-openHPI-MOOCs

Die pandemiebedingte Umstellung sämtlicher Lehre am HPI ab Sommersemester 2020 führte zu großer Dynamik in der digitalen Lehre. Alle im KI-Labor beteiligten Arbeitsgruppen (s. Liste unten: Bert Arnrich: BA, Jürgen Döllner: JD, Tobias Friedrich: TF, Robert Hirschfeld: RH, Christoph Lippert: CL, Christoph Meinel: CM) haben einschlägige Lehrangebote voll digital umgestellt, die eine Aufzeichnung und eine für das digitale Format geeignete Struktur besitzen. Damit wurde erstmalig über mehrere Semester hinweg das KI-relevante Lehrangebot digital verfügbar. Die Kurse bildeten auch die Grundlage für daraus ableitbare MOOCs und openHPI-Kurse.

### Ausgewählte digitale Lehrangebote des KI-Labors:

AP	Titel (Kurs bzw. MOOC)	KI-Labor-Bezüge	SWS	Semester
BA	Statistics in Connected Healthcare	Grundlagen für KI-basierte Methoden und Verfahren	4	SoSe 2020 SoSe 2021
BA	Digital Health für Einsteiger (openHPI MOOC)	Einsatz digitaler Technologien in der Medizin		WiSe 2020 / 2021
JD	3D-Computergrafik II	Grundlagen für GPU-basierte Implementierung und Parallelisierung	4	SoSe 2020
JD	Einführung in die Programmier technik II	Grundlagen der Programmiersprache C++ zur Konstruktion komplexer, performanter Softwaresysteme	4	SoSe 2020
JD	3D-Computergrafik I	Grundlagen für GPU-basierte Implementierung und Parallelisierung	4	WiSe 2020 / 2021
JD	Analysis and Visualization of Spatio-Temporal Data	KI-Verfahren und Methoden für spezifische Domänen	4	WiSe 2020 / 2021
JD	Introduction to Data Visualization	KI-basierte Verfahren der Datenvisualisierung, z.B. hochdimensionale Topic Maps	4	WiSe 2020 / 2021
JD	Systems Engineering and	Programmiertechnische Grundlage für die	4	SoSe 2021

	Data Processing with C++	Konstruktion zeit- und ressourcenkritischer KI-Systeme, -Bibliotheken und -Services		
JD	Software Mining and Applications	KI-Verfahren auf Basis hochdimensionaler Merkmalsräume, generiert durch Repository Mining	4	SoSe 2021
TF	Generating Realistic SAT Instances	KI-Verfahren und Methoden für spezifische Domänen	4	SoSe 2020
TF	Deep Learning for Combinatorial Optimization	Deep-Learning-Verfahren für die kombinatorische Optimierung	4	SoSe 2021
TF	Competitive Programming with Deep Learning	Effiziente Implementierung und Anwendung von Deep Learning Verfahren	4	WiSe 2020/2021
TF	Computational Learning Theory and Beyond	OpenHPI-Kurs zu Grundlagen und formalen Modellen für Machine Learning und AI	2	SoSe 2020
TF	Theorie der Künstlichen Intelligenz	Bachelor-Kurs zu Grundlagen der Künstlichen Intelligenz	2	WiSe 21/22
CL	Introduction to Deep Learning	Grundlagen für Deep Learning als wichtige Form konkreter KI-Implementierung	4	SoSe 2020
CL	Deep Learning	Weiterführende Konzept für Deep Learning als wichtige Form konkreter KI-Implementierung	4	SoSe 2020
CL	Deep Learning	Einführung in Deep Learning	4	SoSe 2021
CL	Attention in Deep Learning	Fundamental Aspekte des Deep Learning	4	SoSe 2021
RH	Code Repository Mining	KI-Verfahren und Methoden für spezifische Domänen	4	SoSe 2020
RH	Softwaretechnik I	Systematische Erstellung komplexer IT-Systeme	4	SoSe 2021
CM	Practical Applications of Deep Learning	KI-Verfahren und Methoden für spezifische Domänen	4	SoSe 2020
CM	Practical Applications of Deep Learning	Deep-Learning in der Praxis der Anwendungsentwicklung	4	SoSe 2021
CM	Machine Intelligence with Deep Learning	Machine Intelligence with Deep Learning	4	WiSe 2021/22

## II Wichtigste Positionen des zahlenmäßigen Nachweises

Im Rahmen der Antragstellung wurde eine genaue Arbeitspaket- und Kostenplanung erstellt. Während der Vorhabendurchführung konnte die Planung der Arbeitspaketstruktur und der Meilensteinplanung eingehalten werden. Die Verteilung der entstandenen Kosten auf die unterschiedlichen Ausgabenpositionen pro Projektpartner wurde im zahlenmäßigen Abschlussbereich aufgeführt und eingehalten. Die Deep Learning Plattform, Speicherserver, SAN/LAN Switches, ML/DL-Daten und -Modelle, Prozessortechnik, Laborsteuerrechner wurden plangemäss beschafft (0850). Die Gesamtfördersumme betrug 2.091.836,00 Euro.

Besondere Abweichungen gab es nicht; fachlich einschlägige Abweichungen wurden durch Umwidmungsanträge genehmigt:

- Umwidmung von Mitteln aus 0843 auf Geräte in 0831, um verschiedene KI-Labor-spezifische Geräte ergänzend zu beschaffen, die sicherstellten, dass die bislang gewonnenen Erkenntnisse sowie die zu diesem Zeitpunkt ausstehenden Arbeiten auf dem aktuellen Stand der technischen Entwicklung durchgeführt werden können.
- Umwidmung von Mitteln aus 0843 auf Geräte in 0850, um die sprunghaften Innovationen im Hardware-Markt (u.a. Sensor-Hardware, Prozessor-Hardware, GPU-Hardware) im KI-Labor angemessen widerspiegeln.

## III Notwendigkeit und Angemessenheit der Projektarbeiten

Der Arbeitsplan des Vorhabens gliederte sich thematisch in arbeitsgruppen-orientierte Arbeitspakete. Ihre Inhalte waren so konzipiert worden, dass sie die zentralen Schlüsseltechnologien für KI-Systemtechnik in Hinblick auf Methodik, Technik und Ausbildung in der IT-Systemtechnik enthielten. Um diese Pakete in ausreichender Tiefe und ausreichendem Umfang bearbeiten zu können und die Ergebnisse im Sinne der förderpolitischen Ziele einsetzen zu können, war die von den Mitarbeitenden geleistete Arbeit vollumfänglich notwendig und angemessen. Diese Arbeit umfasste z.B. die wissenschaftliche Bearbeitung der identifizierten Fragestellungen oder auch die Einbeziehung der Thematiken und Forschungsergebnisse in die Lehre an der Universität Potsdam sowie am Hasso-Plattner-Institut in den nachgelagerten Studiengängen und Online-Kurs-Angeboten.

Konkret hatte das Vorhaben eine Anzahl wissenschaftlich-technischer und wirtschaftlicher Risiken. Bis dato gab es keine in der Praxis große Übersicht über Methodik, Technik und Ausbildungsinhalte der KI-Systemtechnik. Durch das Vorhaben wurde erstmalig eine solche Bearbeitung und Aufstellung möglich, welche die notwendigen inhaltlichen Merkmale abdeckt. Diese hatten für sich jeweils einzeln betrachtet bereits hohe konzeptionelle und technische Herausforderungen, d.h. die Implementierungskomplexität der Einzelverfahren ist hoch; die Komplexität des Gesamtverfahrens steigt jedoch exponentiell. Die Arbeiten im Rahmen des Vorhabens haben sowohl die Einzelverfahrenskomplexität wie auch die Komplexität der Integration und die vielfältigen Wechselwirkungen zwischen den algorithmischen Verfahren in Angriff genommen.

## IV Voraussichtlicher Nutzen/Verwertungsplan

Im Rahmen des Vorhabens wurden die Kompetenzen und das Wissen auf dem Gebiet der KI-Systemtechnik auf Grundlage eines zugeordneten Qualifizierungskonzepts ausgebaut. Das KI-Labor fokussierte dabei insbesondere auf Folgendes:

- Qualifikation von Studierenden: KI als mittlerweile grundlegender Bestandteil von IT-Systemtechnik wurde in Bachelor- und Masterprojekte intensiv eingeführt; die Praxisorientierung von Projekten konnte dadurch massiv gesteigert werden.
- KI-Vertiefungsmodule in allen Studiengängen konnten grundlegend aufgebaut werden und stehen nachhaltig digital bereit (MOOCs, openHPI).
- KI-bezogene Qualifizierung von wissenschaftlichen Mitarbeitenden, u.a. mit dem Ziel der Promotion; hier konnte das KI-Labor eine tragende Rolle übernehmen und zugleich eine inhaltliche Fokussierung liefern. Die Fülle der hochrangigen Publikationen gibt einen Überblick.
- KI-spezifische Hardware und Software: Essentiell für eine Beschäftigung auf dem Stand der Technik und Entwicklung ist das Arbeiten mit hochaktuellen Hardware, Software und Trainingsdaten/-modellen. Auch hier konnte mit dem KI-Labor ein grundlegender Beitrag geleistet werden, der für die nächsten Jahre sehr gut nachgenutzt werden kann, vor allem für wissenschaftlich nachfolgende Arbeiten (Masterarbeiten, Bachelorprojekte und Promotionen).

Nach Abschluss des Vorhabens wurde bereits mit einer Verwertung im wissenschaftlichen Bereich begonnen. Die Ergebnisse des Vorhabens bieten vielfältige wissenschaftliche Anschlussfähigkeiten wie Veröffentlichungen und/oder Präsentationen auf wissenschaftlichen Konferenzen. Die Ergebnisse des Vorhabens begründen damit langfristig das Bearbeiten von Fragestellungen in der KI-Systemtechnik in der Forschung und Entwicklung. Weiterhin wurden mit dem Vorhaben die Grundlagen erzeugt, fortlaufend seitens des Hasso-Plattner-Instituts Abschlussarbeiten und Promotionsthemen zu identifizieren, auszuschreiben und durchzuführen. Mittel- und langfristig bedeutet dies, dass die Universität und das HPI seine Stellung als weltweit sichtbare Größe im Bereich KI-Systemtechnik ausbauen können und damit auch insgesamt der Technologie-Standort Potsdam wesentlich gestärkt wird. Die Verknüpfung hoch spezialisierter, international operierender IT-Unternehmen und Forschungseinrichtungen am Standort Potsdam/Berlin trägt entscheidend zum Ausbau dieser Zukunftstechnologie in der Region bei.

Bereits im Verlauf des Vorhabens hat sich gezeigt, dass mit diesen Aktivitäten ein Beitrag zur Nachwuchsgewinnung geleistet werden kann. Das Vorhaben ermöglichte die Durchführung vielfältiger Forschungsaufgaben und parallel dazu die wissenschaftliche Qualifizierung von KI-Nachwuchskräften.

Der erfolgreiche Abschluss des Vorhabens ermöglicht es dem HPI zukünftig Marktpotenziale und -felder zu erschließen, die nur über eine solche eigenständige, gezielte Ausweitung der strategischen KI-Systemtechnik-Forschung und -entwicklung machbar ist. Weiterführungsmöglichkeiten werden derzeit intensiv auf Landesebene Brandenburg geführt.

Der Knowhow-Transfer wurde durch das KI-Labor insbesondere sichergestellt durch:

- OpenHPI-Kurse, die öffentlich zugänglich sind

- HPI-interne digitale Kurse für die unterschiedlichen Bachelor- und Masterstudiengänge
- Fachverbände: Insbesondere mit dem Verband GeoKOMM wurde ein Fördernetzwerk riskI aufgebaut, das 2022 seine Aktivität aufnahm.

## V Fortschritt auf dem Gebiet

Alle Arbeitsgruppen im KI-Labor haben während der Durchführung des Vorhabens den Fortschritt auf den Forschungs- und Entwicklungsgebieten im gesamten Projektzeitraum stetig verfolgt, überprüft und analysiert. KI als hoch dynamisches Gebiet stellt dabei besondere Herausforderungen. Während der gesamten Durchführungszeit des Vorhabens sind den Arbeitsgruppen keine neuen FE-Ergebnisse von dritter Seite bekannt geworden, die Änderungen des Vorhabens veranlasst hätten. Der Stand der Forschung und Entwicklung und allgemein Fortschritte auf dem Gebiet an anderen Stellen wurden regelmäßig fortgeschrieben. Insbesondere wurden alle einschlägigen Forschungspublikationen und Software-/Produktentwicklungen beobachtet.

Die Herausforderungen wurden sehr gut bewältigt, da im KI-Labor erfahrene Wissenschaftler, überwiegend mit dem Ziel der Promotion oder der Post-Doc-Qualifikation, eigenständig den technisch-wissenschaftlichen Stand des jeweiligen Arbeitsgebiets genau verfolgten. Die Güte der Veröffentlichungen – überwiegend in Peer-Review, hochrangigen Konferenzen und Journalen – stellte zugleich sicher, dass der technisch-wissenschaftliche Stand in Technik, Entwicklung und Wissenschaft geprüft widerspiegelt ist.

## VI Veröffentlichungen

Die in den Arbeitspaketen entstandenen Erkenntnisse, Ergebnisse und Demonstratoren wurden umfänglich in technisch-wissenschaftlichen, einschlägigen Fachzeitschriften, Fachkonferenzen, Workshops, Foren und Journalen fortlaufend veröffentlicht.

**In Abschnitt I dieses Berichts finden sich wichtige ausgewählte Veröffentlichungen pro Arbeitspaket explizit aufgelistet.**

Die Arbeit und die Ergebnisse des KI-Labors konnten so zeitnah und inhaltlich detailliert veröffentlicht werden. Die Publikationen sind im Allgemeinen zugleich elektronisch als PDF über die Web-Seiten der Autoren direkt zugänglich (open access).



