SelfAdapt: Self-supervised Domain Adaptation for Sensor Data

Contact: Prof. Marc Langheinrich
Co-supervisor: Dr. Martin Gjoreski

Wearable devices, combined with Artificial Intelligence (AI) methods, can bring significant and sustainable improvements to our lives – from improved patient monitoring and decreased healthcare costs to enhanced sports performance and improved quality of life. Standard approaches involve Machine Learning (ML) techniques applied to the data captured from body-worn sensing devices. The ML techniques can be based on classical (feature-based) ML, or Deep Learning (DL) applied on the raw sensor data (end-to-end learning). A typical weakness that all ML-based HAR systems have, regardless of whether they are classical or DL-based, is the domain shift that can be caused by different sensor placements. This project will explore personalization and domain-adaptation techniques to address important challenges in wearable computing: noisy data, limited data, and domain shifts in the labels and the sensor data due to subjectivity. ML processing pipelines (including deep learning techniques) will be augmented with the latest unsupervised and self-supervised learning techniques, including contrastive learning. These advanced techniques should produce more robust and data-efficient models (i.e., requiring fewer person-specific labels). Diffusion-based approaches, could also be considered. Project tasks: 1. Overview of existing self-supervised learning approaches; (ii) Pre-process one dataset from wearable sensing systems. Example datasets include emotion recognition, activity recognition and energy expenditure estimation; (iii) Build baseline ML models using the dataset from step (ii); (iv) Develop self-supervised ML approach and compare self-supervised models with the baseline ML models from step (iii).

Physio-RECALL: Analysis of human memory and physiological signals

Contact: Prof. Marc Langheinrich
Co-supervisor: Matias Laporte

Memory Augmentation Systems could be capable of selecting specific to-be-remembered events during an experience, e.g., by detecting the individual as distracted or unengaged. Such information then could be used to generate memory cues for the specific periods during which the user was distracted. One way used to detect the cognitive state of users, e.g. if the user is focused on a task, is through physiological signals, like electrodermal activity (EDA) and interbeat interval (IBI). The goal of this project is to evaluate the potential of physiological signals captured by a wrist-worn device to detect the cognitive load of individuals presented with a memory task. The final purpose is to use this work as an additional input in future Memory Augmentation Systems. The specific tasks of the project are: (i) Overview methods for physiological signal processing and review memory augmentation literature; (ii) Develop the necessary code for processing the EDA and IBI data; (iii) Design and run an experiment to test the setup and pipeline.

Media-RECALL: Analysis of human memory and audiovisual signals

Contact: Prof. Marc Langheinrich
Co-supervisor: Matias Laporte

Memory Augmentation Systems could be capable of selecting specific to-be-remembered events during an experience, e.g., by detecting the individual as distracted or unengaged. Such information then could be used to generate memory cues for the specific periods during which the user was distracted. One way used to detect the emotional state of users, e.g. if the user is engaged on a task, is through audiovisual signals, like speech and facial expressions. The goal of this project is to evaluate the potential of audiovisual signals captured by recording devices (such as microphone, cameras) to detect

the cognitive load of individuals presented with a memory task. The final purpose is to use this work as an additional input in future Memory Augmentation Systems. The specific tasks of the project are: (i) Overview methods for audiovisual signal processing and review memory augmentation literature; (ii) Develop the necessary code for processing the audio and video data; (iii) Design and run an experiment to test the setup and pipeline.

MultiFed: Multimodal Federated Learning for Sensor Data

Contact: Prof. Marc Langheinrich
Co-supervisor: Dr. Martin Gjoreski

Federated learning and its combination with differential privacy is the latest technique for building privacy-aware machine-learning models. Its primary assumption – no data leaves the local data storage, has enabled its application in a variety of privacy-sensitive domains: mobile keyboard prediction, human mobility modeling based on GPS data, modeling from electronic health records, etc. This project will investigate single modality vs. multi-modality federated models. This is an important issue for wearable sensing systems that utilize multiple sensing devices, e.g., smartphone and smartwatch. Each device, and each sensor in the devices, may have a different availability — data coming from the smartwatch may be unavailable at certain periods (e.g., while charging). To enable the collaborative learning of joint models between users with a variable data/modality availability, we will investigate several multi-modal schemes. Project tasks: (i) Pre-process one dataset from wearable sensing systems. Example datasets include emotion recognition, activity recognition and energy expenditure estimation; (ii) Build centralized multimodal and single-modal models using the dataset from step 1; (iii) Build federated multimodal and single-modal models using the dataset from step (i) and compare them with the centralized models from (ii); and (iv) Develop novel multi-modal federated learning method considering device/sensor availability, computational cost, model accuracy.

XAI-Fed: Explainable AI for Federated Models in Wearable Sensing

Contact: Prof. Marc Langheinrich
Co-supervisor: Dr. Martin Gjoreski

Federated learning and its combination with differential privacy is the latest technique for building privacy-aware machine-learning models. Its primary assumption – no data leaves the local data storage, has enabled its application in a variety of privacy-sensitive domains: mobile keyboard prediction, human mobility modeling based on GPS data, modeling from electronic health records, etc. Artificial Intelligence (AI) methods can bring significant and sustainable improvements to our lives. However, end-users must be able to understand those systems. Unfortunately, today's groundbreaking AI methods are black-boxed (i.e., the decision model and the process are not understandable). The increased complexity of AI algorithms has made previous eXplainable AI (XAI) tools unsuitable, including the fact that most of the XAI solutions are not designed to operate under privacy constraints. This project will investigate XAI techniques compatible with privacy-aware approaches (e.g., federated learning). The focus will be on counterfactual explainers [55] for wearable sensing data. Specific project tasks are: (i) Analyze XAI tools that can operate under privacy constraints, focusing on counterfactuals; (ii) Pre-process one dataset from wearable sensing systems. Example datasets include emotion recognition, activity recognition and energy expenditure estimation; (iii) Develop machine learning models for one of the datasets in step 2, and apply existing XAI tools on the models developed, including the method for generating counterfactual explanations, BayCon; (iv) Develop XAI tool for counterfactual explanations that can operate under privacy constraints.

PrivAffect: Privacy-aware personal-video sensing for affect recognition

Contact: Prof. Marc Langheinrich
Co-supervisor: Dr. Martin Gjoreski

Affective computing is an interdisciplinary field that aims at the development of computer science techniques that enable machines to recognize, understand and simulate human affective states. A fundamental assumption is that different mental states (e.g., emotions and stress), and different intensities of those states, manifest through physiological and behavioral changes. A variety of sensing modalities can capture these changes. Video-based sensing is one promising approach for affect recognition, however, it is also privacy intrusive. Thus, this project will develop a method for privacy-aware personal-video sensing for affect recognition. The method will utilize personal camera (e.g., smartphone or laptop camera) and will include privacy-aware features such as: to record only when the users grant permission, to record only when a specific user is in front of the camera (user identification). Once the video is collected in a privacy-aware manner, existing video-based affect recognition software will be sued to extract affect related information. Project tasks: (i) Review existing software (e.g., GitHub and Google scholar) for user identification, software for counting faces in a video, and software for Facial Action Coding System (FACS); (ii) Implement privacy-aware user identification; (iii) Implement privacy-aware method for extracting Facial Action Units (based on FACS); (iv) Test the overall processing pipeline is a small user-study (e.g., 5 to 10 participants).

Fed-CogLoad: Federated Cognitive Load Estimation

Contact: Prof. Marc Langheinrich
Co-supervisor: Dr. Martin Gjoreski

Federated learning (FL) is a state-of-the-art machine-learning technique developed by Google, where the users' privacy is guaranteed by implementing one simple rule: "No personal data leaves the user-device". This project will investigate FL techniques for cognitive load estimate. Cognitive load can be estimated through the analysis of from pupillometry data, brain activation data (EEG), breathing rate, heart rate, heart rate variability and other related physiological responses. In the project you will: (i) overview excising datasets for cognitive load estimation; (ii) develop a centralized machine learning pipeline for cognitive load estimation; (iii) develop a FL pipeline for cognitive load estimation, (iv) compare the centralized and the FL pipeline and write a short report.

Expression Tutor Study

Motivation
One of the most prevalent features of many programming languages is the concept of an **expression**. Expressions occur in many different contexts, such as the right-hand side of assignments, in the argument list of function calls, or in conditions of conditionals or loops. Expressions can be evaluated to produce a value. In statically typed programming languages, expressions have a statically determined type.

Despite the prevalence of expressions in programs, expressions are neglected when teaching programming. At Luce we developed Expression Tutor (https://expressiontutor.org/) to help improve the teaching of expressions. Expression Tutor represents **expressions as trees**, similar to the abstract syntax trees used as internal representations in compilers, and allows students to interactively click together expression trees corresponding to given code snippets. Expression Tutor can automatically generate expression trees for some languages (currently Java and Python), and can automatically correct the trees generated by students.

Goals
Expression Tutor has been used in various kinds of programming courses. While it is visually attractive and seems helpful, it is still unclear how effective it is in helping students to better understand expressions.

In this project, we want to conduct an empirical study to find out. This study will take place in the context of a Python programming summer camp. The study will include quantitative (randomized controlled trial with pre- and post-test) as well as qualitative (interviews, think-alouds) components. You will help to organize and run the study, with a control group, which is taught without any Expression Tutor activities, and an experimental group, which is taught with Expression Tutor activities, and you will help to analyze the results.

Prerequisites
This project requires:

* A thorough understanding of programming language concepts such as expressions, values, types, operators, literals, variables, and functions.
* Strong Python programming skills.
* Strong organizational skills.

More Information
If you are interested in this project, please contact Matthias.Hauswirth@usi.ch to discuss the details.

Web-Based Compositional Graphics Program Construction Kit

Motivation
PyTamaro is an educational Python graphics library for learning to program. The goal of PyTamaro is to instill good programming practices in novice programmers in a motivating way. PyTamaro is easy to learn: Students need only learn a small well-designed API to be able to compose visually attractive complex graphics. PyTamaro consists of functions to create primitive graphics and functions to compose graphics into more complex graphics. PyTamaro is designed in a pure functional style: its types are immutable and its functions are pure. This allows novices to begin programming without worrying about the complex concept of mutation. The coordinate-free design of PyTamaro enables the decomposition of a problem (e.g., draw a clock) into simpler independent subproblems (draw the hands and the clock-face). PyTamaro is a publicly available Python library that can be installed with `pip install pytamaro`. To simplify the use of PyTamaro in classrooms, and to avoid the need to install Python and an IDE, PyTamaro Web (https://pytamaro.si.usi.ch/) provides an in-browser environment for writing PyTamaro code. PyTamaro has been used for training high school informatics teachers in Switzerland, and some of those teachers already adopted PyTamaro in their classrooms.

Goals
Anecdotal evidence from students using PyTamaro indicates that PyTamaro activities are highly motivating. Initial experience shows that the pure functional design of PyTamaro has the potential to lead to cleaner student code, but that the code for constructing more complex graphics is not modularized very well. The student code tends to look like throw-away "scripting" code written for one-time use. In this project we want to develop features for PyTamaro Web that encourage students to write modular, reusable, well-documented, and well-tested code.

Our aim, like with PyTamaro itself, is to exploit the students' intrinsic movitatation: PyTamaro tends to make students write a lot of graphics-producing Python code because they want to produce graphics, not because they want to gain points or a good grade. We want to similarly cause students to write modular, reusable, well-documented, and well-tested code, not because of points or grades, but because of the intrinsic benefits.

Prerequisites
In this project you will develop an extension to our existing PyTamaro Web platform, which is built using Next.JS, React, and Material UI. Therefore, this project requires:

* Strong programming skills in JavaScript, in at least one functional language (such as the Racket student languages, Scala, or Haskell), and in Python.
* Experience with React.
* The ability to understand and resolve problems across complicated technology stacks.

More Information
If you are interested in this project, please contact Matthias.Hauswirth@usi.ch to discuss the details.

FOREWORD

The Software Systems (SWYSTEMS) research group led by Prof. Patrick Eugster provides many opportunities for prospective students to get involved in research, including through internships.

Thanks to generous funding notably from the Swiss National Science Foundation (SNSF) as well as from corporate partners including Cisco, Meta, and SAP, the group supports paid internships year-round, also beyond institutional frameworks such as UROP. So while we encourage the participation in such programs, these are not the only way to start interacting with us.

Do not hesitate to reach out to us to hear more about opportunities. We are always more than happy to talk about our latest research and possibilities for you to contribute. The following list only presents a subset of ongoing projects. We are also happy to talk about possibilities of contributing to smaller tasks within those, or to hear about any project ideas that you may have and to discuss ways in which we may be able to support you.

P. Eugster

Partially Homomorphic Encryption for Stream Processing Frameworks

Supervisors:
- Shamiek Mangipudi <shamiek.mangipudi@usi.ch>
- Dr. Savvas Savvides <savvas@purdue.edu>
- Prof. Patrick Eugster <eugstp@usi.ch>

Due to the rapid spread of IoT, billions of devices are expected to continuously collect and process sensitive data. Because of the limited computational and storage capacity available on IoT devices, the current de facto model for building IoT applications is to send the gathered data to the cloud for computation. Unfortunately, using public (untrusted) cloud infrastructures for processing continuous queries including on sensitive data leads to strong concerns over data confidentiality.

An attractive approach to preserving the confidentiality of continuous query processing while utilizing public clouds is through the use of *partially homomorphic encryption* (PHE). PHE allows computations over encrypted data, without revealing plaintext values, therefore protecting sensitive information. The downside of using PHE is that certain asymmetric PHE cryptosystems are computationally expensive. Recently, we developed a set of *symmetric* cryptosystems that retain the homomorphic expressiveness of previous asymmetric cryptosystems while being more performant (https://github.com/ssavvides/symmetria). The goal of this project is to apply this set of existing symmetric PHE schemes to allow confidentiality-preserving *continuous* queries, using the Apache Storm stream processing framework into which we previously already integrated several PHE schemes (or using an alternative system such as Apache Kafka, Apache Spark Streaming, Apache Flink)

The selected student will take part in 1. exploiting existing support for *user defined functions* (UDFs) in a stream processing framework to integrate the aforementioned symmetric PHE schemes to perform various operations over encrypted data, and 2. perform evaluations to assess the overhead of computations due to the use of PHE. Experience with using a stream processing framework is an advantage. Knowledge of cryptographic constructs and PHE is not required but is considered a plus.

Programmable Elasticity for Stateful Cloud Applications in Orleans

Supervisors:
- Dr. Bo Sang <bsang@purdue.edu>
- Prof. Patrick Eugster <eugstp@usi.ch>

Serverless computing platforms (e.g., AWS Lambda) are among the most popular solutions to manage applications on cloud platforms. They allow developers to program elastic cloud applications consisting of stateless functions that can be automatically scaled in and out, without manual intervention. However, many cloud applications are stateful—while executing, the state of one function needs to be shared with others. Providing elasticity for such stateful functions is much more challenging, as a deployment/elasticity decision for a stateful entity can strongly affect others in ways which are hard to grasp without any application knowledge.

In a paper that appeared at a prime venue, the EuroSys 2020 conference [1], we proposed the PLASMA framework (Programmable Elasticity for Stateful Serverless Computing Applications) to remedy the lack of solution scaling stateful serverless applications. PLASMA includes an elasticity programming language to describe elasticity behavior, and a novel semantics-aware elasticity management runtime that tracks program execution and acts upon application features as suggested by elasticity behavior. The current implementation of PLASMA only supports applications written in AEON [2], an actor-based programming language for actor scalability and event serializability.

The selected student will take part in (1) making PLASMA compatible for applications written in the Orleans [3] actor-based programming language, (2) selecting and/or implementing relevant Orleans applications, and (3) evaluating the effect of PLASMA on these Orleans applications. We plan to use this work as part of an extended version of the EuroSys 2020 paper to be submitted to a journal. Experience with actor-programming models is a plus but is not required.

[1] PLASMA: Programmable Elasticity for Stateful Cloud Applications. Bo Sang, Pierre-Louis Roman, Patrick Eugster, Hui Lu, Srivatsan Ravi, Gustavo Petri. EuroSys 2020 (feel free to ask for a copy of the article)
[2] Programming Scalable Cloud Services with AEON. Bo Sang, Gustavo Petri, Masoud Saeida Ardekani, Srivatsan Ravi, Patrick Eugster. Middleware 2016 (feel free to ask for a copy of the article)
[3] Orleans: Distributed virtual actors for Programmability and Scalability. Philip A. Bernstein, Sergey Bykov, Alan Geller, Gabriel Kliot, Jorgen Thelin. Microsoft Research, Technical report, 2014

Programming Language for Reactive Distributed Monitoring

Supervisors:
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Existing network monitoring solutions primarily focus on querying information from the network rather than on the responses these queries would ensue. Such approaches are limited both in presenting an incomplete abstraction, and in not taking advantage of the many optimization opportunities a holistic view on monitoring and management cycle could bring.

To overcome the limitations of the existing systems we focus on the design that combines monitoring and *management* in a single system. Our system aims at exploiting programmability of network devices to perform as many of the management actions as possible exactly at the point of data collection, i.e., at the switches, thus improving the management reaction time and saving the network capacity otherwise spent on collected data. As not all management decisions can be made locally, we introduce a *distributed* abstraction, where each monitoring and management task is represented by a set of interacting agents spread among several network devices. We call these agents *seeds*. When actually programming these seeds to perform the desired task, several performance factors must be taken into account: 1) switch resources are limited and some are already used by network control plane; 2) polling the data from the switch is a major bottleneck and must be under the system's control; 3) as many monitoring and management tasks can be active in the network simultaneously, the system have a good understanding of seed resource constraints so to utilize resources in the most efficient manner. To relieve the programmer from a burden of low-level resource management, we have designed a domain-specific programming language called *Almanac* and optimization algorithm for resource allocation and seed placement. Almanac features trigger variables for polling and timing, state-machine and messaging abstractions to simplify reasoning of a distributed system, and primitives for expressing placement requirements. At the same time the optimization algorithm takes into account the many resource and placement constraints from the seeds and from the switches to optimize overall utility of the monitoring and management tasks.

The selected student will work on automated translation from *Almanac* to the machine code accessing the existing API of our system, gaining experience in programming language design and implementation. The translation would also need to include simple static analysis to produce the set of optimization constraints for the seed placement algorithm. The exact implementation framework is up to a discussion, but [1], [2], and [3] provide some examples.

[1] https://arzg.github.io/lang/
[2] https://www.stephendiehl.com/llvm/
[3] https://docs.racket-lang.org/guide/languages.html

Cost-based Mechanism Selection for Secure Cloud Computing

Supervisors:
- Shamiek Mangipudi <mangish@usi.ch>
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

The value of what can be derived from customer data is being increasingly recognized by many industries, information is becoming the new currency. At the same, the amount of data generated has been growing exponentially, and many organizations have turned to the cloud in their search for cost-effective information processing. On the account of that, there is a huge demand for processing of sensitive data using third-party untrusted computational resources. While there are both software (homomorphic encryption) and hardware (secure enclaves) techniques with the potential to perform such processing without leaking any information, they have their own constraints and overheads, so that there is no single universal solution. For the best performance different techniques must be combined, but it quickly becomes hard to reason about end-to-end security for non-experts, which data analysts writing queries usually are not.

We have designed a system, called Hydra, that supports a multitude of security mechanisms nicely decoupling privacy policies from the business logic of the queries. To guarantee compliance with a chosen privacy policy, Hydra introduces a lambda-calculus-based domain specific language (DSL), equipped with a type system which ensures the absence of insecure information flows in the system — especially valuable with data propagating through different heterogeneous components/security mechanisms. Hydra system is integrated with the Apache Spark streaming processor to provide users with the familiar query abstraction. The current limitation of Hydra is that the choice of the security mechanism is hardcoded, and while DSL is able to check compliance of any such choice, the flexibility is not fully exploited.

In the course of internship, the selected student will be working on the query optimization part of Hydra (which uses standard Spark SQL extension points). The goal will be to use empirical performance measurements for different security mechanisms combined with Spark execution metrics in order to guide the mechanism choice at the query transformation phase in a fully automated fashion.

Language-based Policy Checking for Secure Computing

Supervisors:
- Shamiek Mangipudi <mangish@usi.ch>
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

With information becoming the new currency, the value of what can be derived from customer data is being increasingly recognized by many industries. At the same, the amount of data generated has been growing exponentially, and many enterprises have turned to the cloud-computing paradigm in their search for cost-effective data processing triggering a huge demand for processing of sensitive data using third-party untrusted computational resources, .e.g., public cloud. Given the inherent limitations and trust issues inhibiting the massive usage of public cloud for sensitive computations, various security mechanisms — both hardware and software — are used to endow the public cloud with the support for privacy-preserving computations by guaranteeing no leakage of sensitive information to the cloud.

We have designed a system, called Hydra, that supports a multitude of security mechanisms while clearly separating the security policies from the business logic of the queries. A security policy checker in Hydra statically checks compliance of queries with the security policy of interest — in particular, ensuring that no insecure information flows exist in the system especially with the involvement of different security mechanisms. As part of the project, the student would undertake the exercise to integrate, enrich and optimize the security policy checker with SparkSQL — the underlying distributed query processing pipeline of Hydra. Our system Hydra is based on Apache Spark — one of the most active open source projects boasting contributions from over 1200 developers spread across 300 companies, making it a unique code base to learn and experience the length and breadth of system design principles.

In the course of the project, the student will gain experience in the implementation of type checking for a programming language augmented with advanced features for information flow tracking, and in the query transformation engine of Spark.

Compile-time Verification of Fault-tolerant Distributed Systems

Supervisors:
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Software defects cost our IT-centered society exorbitant amounts of money. To make matters worse, driven by the advent of paradigms such as cloud computing, blockchains, and the Internet of things, software has been becoming increasingly *distributed*, i.e., its execution spans many processes. Besides having to avoid "conventional" intrinsic defects in the actual software, programmers now have to cater for *partial failures*, e.g., the possibility that certain processes or hosts fail while others continue to operate. Catering for these requires complex protocols, leading to highly error-prone code. Traditional "full-depth" verification of programs involve lengthy verification processes requiring much manual effort and expert knowledge, and are thus easily left out of the loop.

We have adapted a very recent technique for partial software verification, so-called *session types*, to real-life distributed systems, enabling the verification of fundamental properties in the interaction of distributed components (e.g., absence of deadlocks) in a lightweight fashion integrated with compilation of programs [1]. Session types are a form of *behavioral typing*, which, as the latter name suggests, capture behavioral properties of program code including ordering of operations for interaction between distinct components/processes.

Our session typing discipline is based on an *event-driven* programming model widely used in distributed systems, and has been implemented through a domain-specific language (DSL) in Scala [2]. Distributed software systems built using our DSL can thus be easily verified now for salient properties. An initial evaluation based on adapting the cluster manager core component of Apache Spark to our DSL shows only moderate performance overheads compared to the unverified vanilla version (<10%) [1].

The goal of this project is to apply our prototype to further distributed middleware systems. The student will thus be 1. familiarizing themselves with our DSL, 2. and with a middleware system, 3. adapting the system to our DSL, and 4. performing simple performance evaluation to assess overhead based on existing standard benchmarks.

Feedback through lessons learned in the process are of additional value, and room exists for proposing practical extensions and runtime optimizations.

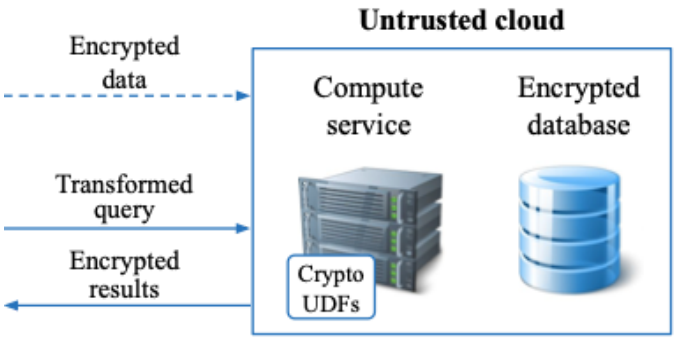
The project will strongly benefit from our prior experience with Spark's cluster manager, and can easily lead to a prolonged research experience, all the way up to a PhD thesis if desired.

[1] Malte Viering, Raymond Hu, Patrick Eugster, Lukasz Ziarek: A multiparty session typing discipline for fault-tolerant event-driven distributed programming. Proc. ACM Program. Lang. 5(OOPSLA): 1-30 (2021).
[2] https://github.com/viering/MPST_toolchain_for_fault-tolerant_distributed_EDP

Privacy-preserving Stream Processing

Supervisors:
- Shamiek Mangipudi <mangish@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Many applications require low-latency processing of streams of data produced at high rates [2]. Such streams often contain private/secret data that must be protected and thus remain encrypted in-flight and in-use. Given the sheer size of stream data, it is ideal to utilize the public cloud for its scalable, cost-efficient infrastructure to process such streams. However, the cloud comes with its own concerns such as shared hardware, multitenancy etc., which adds to the issue of preserving privacy of data in the streams.



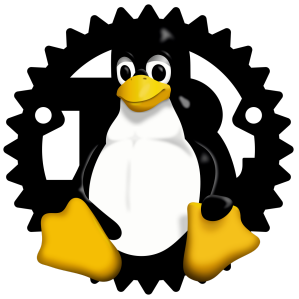
To overcome these concerns, privacy preserving technologies such as partial homomorphic encryption (PHE) have been successfully used to perform operations like filtering, aggregation etc., on encrypted streams without having to decrypt the data. State-of-the-art PHE schemes come with substantial overheads and many techniques have been introduced for reducing the overhead. Recently, efficient practical symmetric PHE (SPHE) schemes have been proposed to reduce the overhead of PHE in batch data processing.

The goal of this project is to investigate the use of symmetric homomorphic encryption schemes for secure processing of encrypted streams – building upon past experience from application of PHE to clouds [2, 3]. Tasks include

- Design of runtime query optimization techniques
- Implementation and evaluation of program analysis for application of SPHE.

[1] S. Savvides, D. Khandewal, and P. Eugster. Efficient Confidentiality-Preserving Data Analytics over Symmetrically Encrypted Datasets. VLDB 2020.
[2] S. Savvides, S. Kumar, J. Stephen, and P. Eugster. C3PO: Cloud-based Confidentiality-preserving Continuous Query Processing. ACM TOPS 2021.
[3] J. J. Stephen, S. Savvides, V. Sundaram, M. S. Ardekani, and P. Eugster. STYX: Stream Processing with Trustworthy Cloud-based Execution. ACM SOCC 2016

Rust for Kernel-level Distributed Services



Supervisors:
- Davide Rovelli <roveld@usi.ch>
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Since its conception more than 50 years ago, C has been a widely adopted programming language in a variety of applications. Today, C is still the de-facto standard for low-level programs, lying at the core of operating systems, device drivers and network protocols. While this ensures good performance, as C can leverage on decades of compiler optimization, it also introduces several security vulnerabilities. This is due to the fact that some of the features in C, namely its explicit memory management and undefined behavior, make it difficult to write secure, bug-free software. Several attempts (e.g. anomaly detection) have been employed on top of existing systems to mitigate C vulnerabilities, but fail to provide a complete solution as C-related security flaws still appear to be one of the most common causes of system errors [2] An alternative approach is to tackle the issue at the source by adopting a different programming language for low-level systems.

Rust is a low-level systems programming language focusing on performance, reliability and robustness [1], created by Mozilla Research in 2009 with the goal of providing an equally-fast alternative to C. Rust achieves this through several features that “force” the programmer to write bug-free software at compile time.

On Dec. 2022, the Linux Kernel added support for Rust [3], increasing its relevance in the systems community. This makes Rust an ideal candidate to write new secure and fast kernel-level services for future distributed systems. However, it still needs to be seen whether Rust can be currently used for this purpose considering that the current Rust support in the Linux Kernel is experimental. Furthermore, the complexity and some of the limitations posed by Rust, make it unclear if it can really substitute C as a leading system programming language [4].

In the course of this project, the student will learn about low-level programming and will evaluate to what extent Rust is currently usable for Linux kernel services. This task will go from analyzing the current Rust kernel environment to attempting to build a proof-of-concept distributed service as a kernel module in Rust.

[1] - https://doc.rust-lang.org/stable/book/foreword.html
[2] - https://msrc-blog.microsoft.com/2019/07/16/a-proactive-approach-to-more-secure-code/
[3] - https://thenewstack.io/rust-in-the-linux-kernel/
[4] - https://dl.acm.org/doi/pdf/10.1145/3133850.3133867

Analysis of Proofs of Unsatisfiability for SMT Solvers

Supervisors:
- Rodrigo Otoni <otonir@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Formal verification aims at providing strong guarantees about the behavior of programs and systems. It relies on logic – be it propositional, first-order, or higher-order logic – to precisely describe the program or system in question and check any desired properties, e.g., deadlock-freedom on a concurrent system. Both academia and industry recognise the need for formal verification, since the increasing complexity of applications leads to bugs occurring more frequently. E.g., formal verification is used extensively at Amazon Web Services [1].

Many verification approaches rely on automated reasoning engines for first-order logic. They ultimately reduce the verification problem to checking the satisfiability of first-order logic formulas. Such formulas are commonly expressed in a form suitable to satisfiability modulo theories (SMT) solvers [2], which makes these solvers a critical part of the verification infrastructure. Despite being used to ensure correctness, SMT solvers are known to have bugs. This undermines all guarantees given as to the behavior of a program. To tackle the problem of bugs leading a SMT solver to an incorrect result, artifacts explaining the achieved result can be produced by the solver. In case the logic formula provided as input is satisfiable, i.e., there is an assignment to its variables that makes the formula be evaluated to true, the assignment itself can be such an artifact. If the formula is unsatisfiable, however, there is no standard artifact available. In the last decade, the idea of a proof of unsatisfiability became popular, arising in solvers for Boolean satisfiability [3] and then reaching SMT solvers [4,5].

The goal of this project is to analyze existing proof formats used by different SMT solvers. The student will learn the basics of SMT solving and will endeavor to understand the proof formats used by four SMT solvers – Z3, CVC5, veriT, and OpenSMT. The formats will be compared both quantitatively, via an evaluation using publicly available benchmarks, and qualitatively, via a description of the elements underpinning each proof format.

[1] C. Newcombe et al.. How Amazon Web Services uses Formal Methods. Communications of the ACM 58(4), pp. 66–73 (2015).
[2] C. Barrett and C. Tinelli. Satisfiability Modulo Theories. Handbook of Model Checking, pp. 305-343 (2018).
[3] N. Wetzler et al.. DRAT-trim: Efficient Checking and Trimming Using Expressive Clausal Proofs. 17th Int. Conf. on Theory and Applications of Satisfiability Testing, pp. 422-429 (2014).
[4] C. Barrett et al.. Proofs in Satisfiability Modulo Theories. (2014).
[5] R. Otoni et al.. Theory-Specific Proof Steps Witnessing Correctness of SMT Executions. 58th ACM/IEEE Design Automation Conference, pp. 541-546 (2021).

Leveraging Synchrony for Efficient Distributed Services

Supervisors:
- Davide Rovelli <roveld@usi.ch>
- Dr Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Among a wide range of user services, an ever-growing number of software applications are implemented and deployed as *distributed* systems. Distribution is in fact an essential requirement for cloud-based services that want to ensure high availability and fault tolerance through replication. Common use-cases are applications dealing with large amounts of data such as online databases, streaming services, and shared file-storage.

A core problem of distributed systems is achieving consistency among distributed processes. This requires some form of coordination, a longstanding issue considering the asynchronous behavior of current networks and end hosts derived from the presence of arbitrary delays in the communication and processing times (due to, e.g., network congestion, hardware faults). Several asynchronous algorithms and services have been developed over the years to deliver robust coordination, but do so by sacrificing performance and increasing complexity.

We propose a different approach by tackling the problem at the source using a *synchronous* system to achieve efficient coordination. Together with SAP, we developed a prototype system (X-Lane) that achieves practical synchrony on end-to-end process interaction [1], paving the way for faster, simpler, synchronous distributed datacenter-based services.

During this project, the student will work on the simulation environment for X-Lane on the OMNet++ simulator. The main task will consist in designing/selecting and implementing a distributed algorithm using X-Lane, choosing from a number of well-established primitives such as leader election (part of Raft [2], [3]), or causal total order broadcast ([4] [5]). Two students have already successfully worked on such primitives for their BSc theses.

[1] P. Jahnke et al.. Live in the Express Lane. USENIX Annual Technical Conference. 2021. https://www.usenix.org/system/files/atc21-jahnke.pdf
[2] Diego Ongaro et al.. In search of an understandable consensus algorithm. USENIX Annual Technical Conference. 2014
[3] Marcos Kawazoe Aguilera et al.. Stable Leader Election. International Conference on Distributed Computing (DISC '01). 2001.
[4] Kenneth Birman et al. Lightweight causal and atomic group multicast. ACM Trans. Comput. Syst. 9. 1991
[5] C. Cachin, R. Guerraoui, and L. Rodrigues. Introduction to Reliable and Secure Distributed Programming. Springer Berlin Heidelberg, 2011. Chapter 3.9, page 100.

Embedding Proofs of Unsatisfiability for SMT Solvers into CHC Solving

Supervisors:
- Rodrigo Otoni <otonir@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Formal verification aims at providing strong guarantees about the behaviour of programs and systems. It relies on logic – be it propositional, first-order, or higher-order logic – to precisely describe the program or system in question and check any desired properties, e.g., deadlock-freedom on a concurrent system. Both academia and industry recognise the need for formal verification, since the increasing complexity of applications leads to bugs occurring more frequently. E.g., formal verification is used extensively at Amazon Web Services [1].

Constrained Horn Clauses (CHC) are a fragment of first-order logic that naturally captures many verification problems, including, for instance, the checking of safety properties of concurrent and distributed systems [2]. Reasoning about first-order logic formulas boils down to satisfiability queries, commonly solved by satisfiability modulo theories (SMT) solvers [3]. SMT solving is thus fundamental to CHC-based verification techniques. Despite being used to ensure correctness, SMT solvers are known to have bugs, which undermines all guarantees given as to the behaviour of a program being analysed. To tackle this problem, proofs of unsatisfiability for SMT solvers can be used [4,5]. These proofs are artefacts produced by SMT solvers that validate that a given formula is unsatisfiable, i.e., that there is no assignment to the formula's variables that makes it evaluate to true. Recently, many SMT solvers started to have the option to produce such proofs. Verification tools, however, don't commonly benefit from these proofs to strengthen the guarantees about the final result provided to the user.

The goal of this project is to embed SMT proofs of unsatisfiability into CHC-based verification. The student will learn about the basics of CHC solving and of proofs of unsatisfiability for SMT, and will then integrate the proofs into the pipeline of the Golem CHC solver [6].

[1] C. Newcombe et al.. How Amazon Web Services uses Formal Methods. Communications of the ACM 58(4), pp. 66–73 (2015).
[2] A. Gurfinkel and N. Bjørner. The Science, Art, and Magic of Constrained Horn Clauses. 21st Int. Symposium on Symbolic and Numeric Algorithms for Scientific Computing, pp. 6-10 (2019).
[3] C. Barrett and C. Tinelli. Satisfiability Modulo Theories. Handbook of Model Checking, pp. 305-343 (2018).
[4] C. Barrett et al.. Proofs in Satisfiability Modulo Theories. (2014).
[5] R. Otoni et al.. Theory-Specific Proof Steps Witnessing Correctness of SMT Executions. 58th ACM/IEEE Design Automation Conference, pp. 541-546 (2021).
[6] See https://github.com/usi-verification-and-security/golem.

Validation of Quantum Policies with NetSquid

Supervisors:
- Anita Buckley <buckla@usi.ch>
- Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Quantum computing, communication and sensing technologies offer fundamentally new ways for information processing. The objective of quantum communication is to transmit quantum bits (qubits). Qubits can be entangled, causing stronger correlations over large distances than are possible with classical information. The no-cloning theorem (i.e., qubits cannot be copied) makes quantum communication inherently secure, leading to several novel applications [1]. Quantum networks enable quantum-secure communication and entanglement-assisted communication. Due to entanglement, quantum networks with very modest resources outperform classical communication.

The distribution of entangled qubits (Bell pairs) between distant end-nodes will be the main task of the quantum internet of the future [2]. We are developing a language and logic for dealing with and reasoning about quantum networks, QNetKAT (Quantum NetKAT, inspired by [3]). QNetKAT has primitives for creating and transmitting Bell pairs, together with parallel and sequential composition operators, and offers a simple way for expressing quantum network policies.

In the course of this project the student will get familiar with the components of quantum networks and protocols for long distance entanglement distribution. The main task will consist of designing quantum protocols in the QNetKAT language and implementing them using the NetSquid quantum network simulation platform using Python [4].

[1] S. Pirandola, U. L. Andersen, L. Banchi et al. Advances in Quantum Cryptography. (2022) https://arxiv.org/pdf/1906.01645.pdf
[2] J. Illiano, M. Caleffi, A. Manzalini and A. S. Cacciapuoti. Quantum Internet Protocol Stack: A Comprehensive Survey. https://arxiv.org/pdf/2202.10894.pdf
[3] C. Anderson, N. Foster, A. Guha, J.-B. Jeannin, D. Kozen et al. NetKAT: Semantic Foundations for Networks. ACM SIGPLAN Notices, 49 (1), 113–126 (2014).
[4] T. Coopmans, R. Knegjens, A. Dahlberg et al. NetSquid, a NETwork Simulator for QUantum Information using Discrete events. Commun Phys 4, 164 (2021).

Formal Modeling of Probabilistic Quantum Network Policies

Supervisors:
- Anita Buckley <buckla@usi.ch>
- Pavel Chuprikov <chuprp@usi.ch>
- Prof. Patrick Eugster <eugstp@usi.ch>

Quantum computing, communication and sensing technologies offer fundamentally new ways for information processing. The objective of quantum communication is to transmit quantum bits (qubits). Qubits can be entangled, causing stronger correlations over large distances than are possible with classical information. The no-cloning theorem (i.e., qubits cannot be copied) makes quantum communication inherently secure, leading to several novel applications [1]. Quantum networks enable quantum-secure communication and entanglement-assisted communication. Due to entanglement, quantum networks with very modest resources outperform classical communication.

The distribution of entangled qubits (Bell pairs) between distant end-nodes will be the main task of the quantum internet of the future [2], and the main challenge will be scaling. We are developing a language and logic for dealing with and reasoning about quantum networks, QNetKAT (Quantum NetKAT, inspired by [4]). QNetKAT has primitives for creating and transmitting Bell pairs, together with parallel and sequential composition operators, and offers a simple way for expressing quantum network policies.

In the course of this project the student will get familiar with the components of quantum networks and protocols for long distance entanglement distribution. Decoherence, losses and noise-errors cause stochastic behaviour of quantum operations [3]. The goal of this project is to develop the QNetKAT language with a probabilistic semantics. The main task will consist of extending the language with new primitives for expressing probabilistic behaviours [5].

[1] S. Pirandola, U. L. Andersen, L. Banchi et al. Advances in Quantum Cryptography. (2022) https://arxiv.org/pdf/1906.01645.pdf
[2] J. Illiano, M. Caleffi, A. Manzalini and A. S. Cacciapuoti. Quantum Internet Protocol Stack: A Comprehensive Survey. https://arxiv.org/pdf/2202.10894.pdf
[3] S. Brito, A. Canabarro, R. Chaves, and D. Cavalcanti. Statistical Properties of the Quantum Internet. Phys. Rev. Lett. 124, 210501 (2020).
[4] C. Anderson, N. Foster, A. Guha, J.-B. Jeannin, D. Kozen, C. Schlesinger and D. Walker. NetKAT: Semantic Foundations for Networks. ACM SIGPLAN Notices, 49 (1), 113–126 (2014).
[5] N. Foster, D. Kozen, K. Mamouras, M. Reitblatt and A. Silva. Probabilistic NetKAT. Programming Languages and Systems, ESOP 2016.

Portable Programmer-agnostic use of Trusted Hardware

Supervisors:
- Prof. Patrick Eugster <eugstp@usi.ch>
- Dr Pavel Chuprikov <chuprp@usi.ch>

Malware and other attempts of tampering with computer software remain a dominant challenge to computer security. While several trusted execution environments (TEEs) allowing programs to be shielded from attacks exist (e.g., Intel SGX, ARM Trustzone, AWS Nitro), leveraging these requires expert knowledge in security and respective TEE mechanisms, in addition to deep understanding of the corresponding programs. Even without considering the performance characteristics of different TEE offerings and the different constraints they put on software using them, TEE-based programs are not portable across TEEs of different vendors due to different APIs proposed and functionalities.

We thus propose to use TEEs in combination with *program anomaly detection* (AD) [1]. By creating models of programs and comparing these against the actual executions at runtime, AD can be applied without modifications to software. By tracing the appropriate features, AD has been shown to be able to detect both control-oriented and data-oriented attacks [2] with low overheads [3, 4]. However, being implemented fully in software, existing solutions have a fundamental flaw – their own mechanisms for tracing executions and comparing them to models by the means of monitors are not protected from tampering. We thus envision an architecture where an AD monitor is implemented in a secure manner leveraging a TEE and other hardware features. This generic implementation of the monitor is independent of any monitored program and thus easily portable across TEEs of different vendors; it has to be adapted only once per target platform.

The goal of this project is to select an existing learning-based AD solution, and implement its AD monitor leveraging Intel SGX as first proof-of-concept platform.

This project is funded by Meta through an award for Research in Privacy and Security, and offers ample opportunities for continued involvement. Working in groups is also possible.

[1] Ghosh, A. Schwartzbard, and M. Schatz. Learning Program Behavior Profiles for Intrusion Detecion. Workshop on Intrusion Detection and Network Monitoring, 1–—13, 1999.
[2] L. Cheng, K. Tian, and D. Yao. Orpheus: Enforcing Cyber-Physical Execution Semantics to Defend Against Data-oriented Attacks. ACSAC, 315–—326, 2017.
[3] K. Xu, K. Tian, D. Yao, and B. G. Ryder. A Sharper Sense of Self: Probabilistic Reasoning of Program Behaviors for Anomaly Detection with Context Sensitivity. DSN, 467–478, 2016.
[4] L Cheng. Program Anomaly Detection Against Data-oriented Attacks. Ph.D. dissertation, Virginia Tech, 2018.

Personalization in Conversational Search

Conversational IR (CIR) aims to enable the user to interact with the IR system in a "conversational" manner. In other words, it allows the user to seek information via a multi-turn conversation with the IR system in natural language, in either spoken or written form. The CIR system should be able to understand the user query and interact with the user to understand better the user information need as expressed by the query.

Your task is to gather multiple users, coming from various backgrounds (thus having different knowledge levels and personas - these could be other students, staff, or any person you know). The users would then participate in a Wizard-of-Oz (WoZ) study with a fictitious CIR. In other words, the users would interact with a conversational search interface thinking they interact with an automatic system, while on the other side there is a wizard, in other words, a human (you), pretending to be the system. In that way, we are allowed to explore how different users interact with a conversational system in a controlled environment, as the wizard will have to follow a script and somehow behave like a "machine".

Your task is the following:

- Select a dataset containing appropriate set of information needs and an appropriate collection of documents (e.g., ClariQ - provide to you);
- Gather 10-20 users (for example other students, possibly with different backgrounds) to participate in user study.
- Write guidelines for the Wizard for the WoZ study – the procedure for answering users queries (we can provide papers that describe the process clearly).
- Conduct the WoZ study where users interact with the "system" (actually a Wizard) for multiple conversational turns.
- Find patterns between different groups of users – how do they formulate queries? how do they diverge in subsequent conversational turns?
- Retrieve results from a collection by using users queries (using Python, HPC environment, Anserini)
- Compare and visualize different result rankings (computing Spearman's rank coefficient, statistical significance, visualizations between sets of documents, dissecting and visualizing data by user groups and patterns)
- Optional (if initial study is promising and you have time): scale up the experiment using crowdsourcing (Mturk) to involve more users with an even more variable background.

References:

- Wizard of Oz original paper
  - What is WoZ?
- Wizard of Oz for Conversational Search
  - How to use WoZ for mixed-initiative conversational search?
- Where do queries come from?
  - Given the same information need, users formulate the queries differently, which results in different results presented by the search engine.
  - http://marksanderson.org/publications/my_papers/sigir2022a.pdf

The project will be supervised by Prof. Fabio Crestani and co-supervised by Ivan Sekulić.

DEA - Designing and Educational Agent to Detect and Mitigate Gender Stereotypes in the Classroom

Data on gender distribution in education, employment and career show that Informatics is male dominated. The first barrier encountered by girls on their way to entering the world of Informatics is at school level. Female students often believe they are not good enough, lack in experience and so find Informatics not appealing. This negative attitude is also due to male and female roles being stereotypically defined in society where Informatics is labelled as a male thing. We argue that this bias and its influence on study and career choices of girls and young women have to be addressed if we want society to benefit from inclusive, useful, usable, and attractive technological solutions for all to use.

SNSF funded, project TADAA - **T**ools for **A**ssessing and **D**eveloping **A**ffecting & **A**ttractive Narratives for Girls in Informatics - aims boosting on-going research and providing evidence of the necessary steps to foster a needed shift. The design process will be driven by teachers, children and parents working with researchers. Together, we will define activities and procedures to support inclusive teaching. At the same time, we will study how technology can play a supportive role in detecting evidence and raising awareness on the presence of stereotypes and their influence on children's decisions about their future study and careers.

Here, we propose an exploration into how technology could support teaching Informatics as a means to promote such a change.

Educational Agents have been used for a few years now but the sensational advances in technology and the emergence of AI powered tools such as CHAT-GPT have made this area of research more exciting than ever!

We propose to design and develop an educational agent that will embody educational skills with respect of gender differences and will foster in children the concept of diversity and inclusion.

If you are interested in:

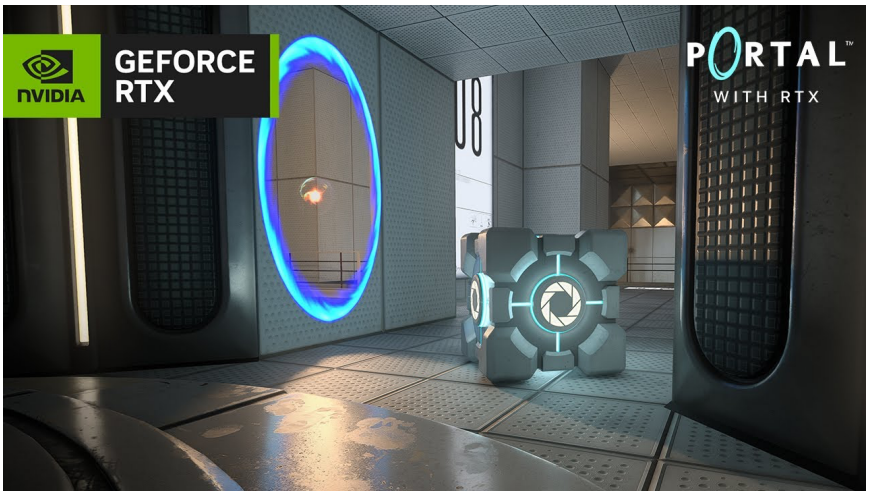
- Learning how to elicit and understand children's preferences regarding educational agents
- Working on the design of a brand-new agents for children, from the aesthetic to the actual functionality.
- Learning how to work with Arduino modules, pixel matrices and other fun technology to create a true interactive experience for children.

This is the perfect project for you! You will work together with a team of researchers, developing your coding and designing skills, and you will participate in a real-world research project with real users and stakeholders!

Reinforcement Learning and Neural Radiance Fields for Real-time Ray Tracing

Advisors: Piotr Didyk, Jorge Condor
Perception, Display, and Fabrication Group (https://www.pdf.inf.usi.ch)

Motivation



Path tracing and ray tracing techniques are the holy grail of computer graphics. They are simple, elegant solutions to the physical problem of light transport, how light flows from its source and bounces around a scene to reach our eyes. However, the nature of the algorithm, namely its immense computational cost, has limited its adoption to high-budget cinematic productions and offline simulations, while real-time applications like videogames had to resort to the cheaper but far less realistic rasterization, where light behavior is "baked", or rather, placed by artists, instead of resembling an actual physical simulation. As of the last few years though, the paradigm is changing: hardware accelerators are being included in GPUs to accelerate some of the costlier parts of the ray tracing algorithm (NVIDIA's RTX cores are just ray-triangle intersection checkers). Thus, the advent of real-time raytracing is upon us, which will dramatically change the way we design GPUs and videogames, delivering unprecedented levels of realism.

Problem Statement

We present a longstanding problem of ray tracing (RT): the rendering of complex luminaires and grand chandeliers. Due to the nature of RT, every change of medium will generate a new bounce, with the goal of all rays (which in RT, come from the camera) being to reach sources of light on the other end. But, the more bounces we have (i.e. several layers of crystal), and the smaller the light sources are (small wires of incandescent bulbs), the smaller the chance of hitting the light source. In practice, this means that scenes containing these super-complex light sources will require a lot more computation to be rendered, and thus they are completely out of reach for real time applications like modern videogames. Our solution to the problem is to abstract the final scene of all the internal complexity of these grand chandeliers by encoding them using Neural Radiance Fields (NeRFs) which allow us to consider them as a single view-dependent source of light. However, knowing where to direct our rays when all we have is a black-box neural network is impossible, and results in many rays being wasted, falling into areas of the luminaire where no light is being emitted (i.e. the brass handles, or empty spaces within the different branches). We propose to use reinforcement learning to train a model that learns the emission profile of the luminaire and is thus able to predict the most energetic paths, enabling what in computer graphics is known as importance sampling (simply, sending rays in the directions that matter).

Your Task

Your task will be to develop and train a model using reinforcement learning so that it learns where most energy is coming from a view-independent octree structure that encodes our NeRFs. Upon completion, we will integrate it within our NeRF-driven path tracing pipeline to significantly accelerate it.

Why You Should Choose This Project

If you have interest in computer graphics or AI, this project combines both in a very practical and potentially impactful way. By the end of this project, you will have learned about state of the art techniques in computer vision (NeRF), computer graphics (path tracing) and AI (reinforcement learning), which are key technologies for the future and some of the main drivers of innovation in top tech companies right now (Google, Microsoft, NVIDIA, Meta, among others). Furthermore, this is a research-oriented project. If successful, we envision it to become a part of a submission to one of the top-tier computer graphics conferences such as ACM SIGGRAPH.

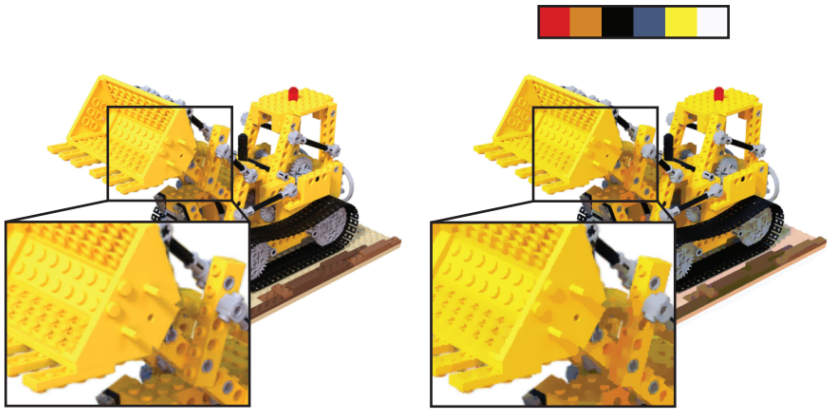
Further Information

Please feel free to contact us for more information at piotr.didyk@usi.ch and jorge.condor@usi.ch.

Perceptual Compression of Neural Radiance Fields

Advisors: Piotr Didyk, Jorge Condor
Perception, Display, and Fabrication Group (https://www.pdf.inf.usi.ch)

Motivation



Neural Radiance Fields or NeRFs for short are one of the most interesting technologies to come up in the past 3 years. They enable impressive reconstructions of real scenes with just a few images, through the use of basic computer graphics algorithms and machine learning. These reconstructions are behind some of the biggest advances in the last few years in computer graphics (real time path tracing), computational photography (next denoising for phone cameras) and even autonomous driving (improved SLAM algorithms), among many others.

Problem Statement

However, reconstructing novel views from this latent space is very expensive, or requires dedicated hardware (new GPUs with large amounts of cache). Nevertheless, there are ways to circumvent this by extracting the latent space into an octree, which is orders of magnitude faster. The side of effect of this is that, while the latent model occupies around 10Mb, an octree with good quality can weight several gigabytes, which is not feasible for most mobile applications.

Your Task

Your task will be to research ways of compressing this octree by means of exploiting the human visual system. For example, due to perceptual effects, humans are more sensitive to certain colors and less to others. By exploiting this it is possible to use less data to map certain areas of the color space, while maintaining the same perceived quality. This trick is normally used in video compression algorithms, but it has not yet been applied to 3D reconstructions of this kind.

Why You Should Choose This Project

If you have interest in computer graphics or AI, this project combines both in a very practical and potentially impactful way. By the end of this project, you will have learned about state of the art techniques in computer vision (NeRF), and understood basic notions of human perception. Furthermore, this is a research-oriented project. If successful, we envision it to become a part of a submission to one of the top-tier computer graphics conferences such as ACM SIGGRAPH.

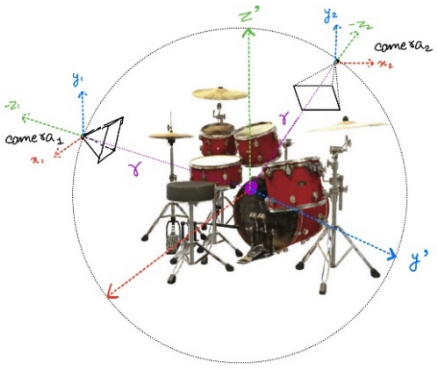
Further Information

Please feel free to contact us for more information at piotr.didyk@usi.ch and jorge.condor@usi.ch.

Complex Camera Models for Neural Radiance Fields

Advisors: Piotr Didyk, Jorge Condor
Perception, Display, and Fabrication Group (https://www.pdf.inf.usi.ch)

Motivation



Neural Radiance Fields or NeRFs for short are one of the most interesting technologies to come up in the past 3 years. They enable impressive reconstructions of real scenes with just a few images, through the use of basic computer graphics algorithms and machine learning. These reconstructions are behind some of the biggest advances in the last few years in computer graphics (real time path tracing), computational photography (next-gen denoising for phone cameras) and even autonomous driving (improved SLAM algorithms), among many others.

Problem Statement

While NeRF works incredibly well with cameras where the pin-hole assumption holds (i.e. phones), it is not capable of modelling more complex camera systems, where blur from de-focused areas created by using big focal distances or very fast lenses can create too much ambiguity for NeRF to work properly. However, unlike pin-hole cameras, where everything is in focus, these blurred areas can provide depth information, which if modelled correctly would notably increase the performance of the reconstruction, both for rendering and for the quality of the reconstruction itself.

Your Task

Your task will be to implement more complex camera models within the NeRF framework, allowing the method to exploit wider apertures, different exposures, focal lengths and obtain depth cues from defocus. Potentially, one could exploit NeRF this way to obtain full light fields from several takes from a single view point at different focus points.

Why You Should Choose This Project

If you have interest in computer graphics or AI, this project combines both in a very practical and potentially impactful way. By the end of this project, you will have learned about state of the art techniques in computer vision (NeRF), and computational photography (complex camera models and the digital processing pipeline). Furthermore, this is a research-oriented project. If successful, we envision it to become a part of a submission to one of the top-tier computer graphics conferences such as ACM SIGGRAPH.

Further Information

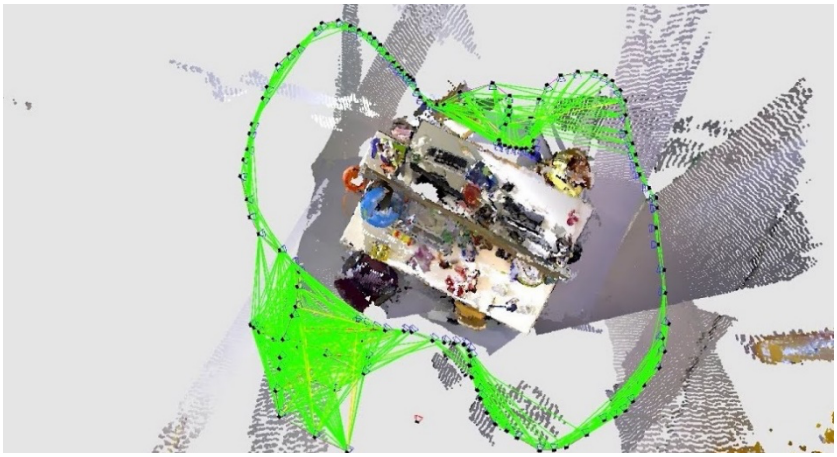
Please feel free to contact us for more information at piotr.didyk@usi.ch and jorge.condor@usi.ch.

Progressive Neural Radiance Fields for Robotic Path-Planning

Advisors: Piotr Didyk, Jorge Condor
Perception, Display, and Fabrication Group (https://www.pdf.inf.usi.ch)

Motivation

Neural Radiance Fields or NeRFs for short are one of the most interesting technologies to come up in the past 3 years. They enable impressive reconstructions of real scenes with just a few images, using basic computer graphics algorithms and machine learning. These reconstructions are behind some of the biggest advances in the last few years in computer graphics (real time path tracing), computational photography (next-gen denoising for phone cameras) and even autonomous driving (improved SLAM algorithms), among many others.



Problem Statement

NeRF reconstructions are notorious for the amount of different viewpoints required to obtain high quality reconstructions; otherwise, uncertainty in the reconstruction manifests mainly in blur and non-smooth surfaces, and most notably in very long decays in the transmittance function. This problem is homonymous in traditional reconstructions based on feature extraction and bundle adjustment, while discovering views with high uncertainty is more expensive. Furthermore, in NeRF, solving the optimization for hundreds of views takes much longer to converge, which makes it particularly interesting to select only those that will make a big difference on reconstruction quality. With the advent of simultaneous localization and mapping (SLAM) algorithms based on NeRF for autonomous driving and NeRF-based UAV 3D mappings, there's an increasing need for a method to assess where uncertainty is highest in a NeRF model, to suggest which paths or views should be taken next to reconstruct a scene with the highest level of quality at the lowest number of views possible.

Your Task

Your task will be to develop a method that takes the uncertainty cues NeRF models provide and suggest which views should be added next to the optimization to maximize quality at the lowest number of views possible. Ideally, a progressive way of training NeRFs could be used or developed, in which case this could have impact in a wider range of applications, particularly those with unconstrained datasets and spaces, reducing the overall training time of NeRFs. If time allows, your method could be deployed and combined with existing robotic path planning systems for showcasing a real application.

Why You Should Choose This Project

If you have interest in computer graphics or AI, this project combines both in a very practical and potentially impactful way. By the end of this project, you will have learned about state-of-the-art techniques in computer vision (NeRF), and computational photography (complex camera models and the digital processing pipeline). Furthermore, this is a research-oriented project. If successful, we envision it to become a part of a submission to one of the top-tier computer graphics conferences such as ACM SIGGRAPH or robotics conferences such as IROS.

Further Information

Please feel free to contact us for more information at piotr.didyk@usi.ch and jorge.condor@usi.ch.

Anisotropic multiresolution analyses for deepfake detection

Contact: Prof. Michael Multerer (multem@usi.ch), Wei Huang (huangw@usi.ch), and Dr. Davide Baroli (davide.baroli@usi.ch)

Generative Adversarial Networks (GANs) and diffusion models have paved the path towards entirely new media generation capabilities at the forefront of image, video, and audio synthesis. However, they can also be misused and abused to fabricate elaborate lies, capable of stirring up public debate. The threat posed by GANs and diffusion models has sparked the need to discern between genuine content and fabricated one. Previous studies have tackled this task by using classical machine learning techniques, such as k-nearest neighbors and eigenfaces, which unfortunately did not prove very effective. Subsequent methods have focused on leveraging frequency decompositions, i.e., discrete cosine transform, wavelets, and wavelet packets, to preprocess the input features for classifiers. During this summer project, you may investigate all kinds of features obtained by multiresolution methods for detecting fake images. Especially we focus on anisotropic features because GANs and diffusion models primarily utilize isotropic convolutions to generate their output, they leave clear traces, their fingerprint, in the coefficient distribution on sub-bands extracted by anisotropic transformations.

We are open-minded when it comes to the concrete design of the project and any idea is welcome. Access to the ICS cluster's GPU resources for your experiments will be provided. You may extend this project to your master thesis and possibly to a scientific publication at a workshop or conference of machine learning.

Prerequisites:

- Knowledge of programming, e.g., Python, Tensorflow or PyTorch.
- Knowledge of fundamental deep learning technique, GANs and diffusion models are preferred.
- Knowledge of wavelet transform is not necessary but highly appreciated.

Which one is real? If you want to know the answer, please do not hesitate to contact us by e-mail for more details.

Reliability Estimation for Deep Learning Systems under Operational Distribution Shift

Paolo Tonella
TAU research group, Software Institute

Abstract

Deep learning systems are trained in lab environment, where train and test datasets are collected in a constrained and artificial setting, which may deviate substantially from the operational, in field setting. Such a discrepancy is called the operational distribution shift and may result in over estimation of the system's reliability. In fact, the reliability measured in the test set, which is still collected in a lab environment, might be higher than the reliability actually experienced in the field.

To address the operational distribution shift, two techniques have been recently proposed in the literature: *adaptive sampling* [1] and *calibrated uncertainty* [2]. The idea behind adaptive sampling is to sample the operational environment with inputs to be labelled manually, in a way that is biased toward system failures. Then, the estimation of reliability based on such sampled data points is adjusted to take the bias into account, resulting in an estimate of reliability which is equivalent to that obtainable by simple random sampling, but provides a richer set of failure examples. The main cost of this approach is manual labelling of the sampled inputs. The second approach takes advantage of a white box uncertainty estimator to obtain a calibrated measure of confidence (or reliability). Calibration does not require any manual labelling, because uncertainty can be estimated at inference time with no knowledge of the ground truth.

The goal of this project is to compare the accuracy of reliability estimation achieved by adaptive sampling vs calibrated uncertainty, in the presence of an operational distribution data shift. The Summer student will: (1) familiarize with both techniques by reading the related papers and replicating the results described in the papers; (2) selecting a benchmark of deep learning systems to which both approaches are applicable; (3) design and execute a comparative empirical study, to determine which technique provides a better approximation of the actual operational system reliability.

Contact: paolo.tonella@usi.ch

References

[1] Antonio Guerriero, Roberto Pietrantuono, and Stefano Russo. Operation is the hardest teacher: estimating DNN accuracy looking for mispredictions. In *43rd IEEE/ACM International Conference on Software Engineering, ICSE 2021, Madrid, Spain, 22-30 May 2021*, pages 348–358, 2021.

[2] Ranganath Krishnan and Omesh Tickoo. Improving model calibration with accuracy versus uncertainty optimization. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Generating Valid Test Inputs for Deep Learning Systems

Summer Project Proposal

Paolo Tonella, Matteo Biagiola
Software Institute@USI
{paolo.tonella|matteo.biagiola}@usi.ch

Deep Learning (DL) models are being used to address a variety of tasks, from image classification to conversational chatbots. As DL models are being deployed in production, it is of paramount importance to test their capabilities.

Several Test Input Generators (TIGs) have been proposed in the literature to test the generalization capabilities of DL models with artificially crafted inputs in order to induce a misclassification.

However, such artificially crafted inputs might not be *valid*, i.e., belonging to the input domain the model has been trained on. On the other hand, automatically checking the validity of the generated inputs is difficult, as validity is hard concept to formalize and, hence, automate.

Project Proposal

The goal of the project is to develop a test generation pipeline for DL models that exploit human feedback to automatically validate generated inputs. The project will focus on image classification as a DL task. An input is valid for a DL model if it is recognizable by human domain experts in the input domain, i.e., an input to which a human can confidently assign a label taken from the input domain.

For instance, let us consider the handwritten digit classification task using the MNIST dataset.

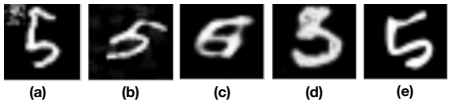


Figure 1: MNIST digits generated by different test generators.

Figure 1 shows the image of a "5" digit generated by different TIGs proposed in the literature. Such inputs are misclassified by the DL model under test, i.e., the label predicted by the model is different than 5. However, the validity, and hence the significance, of some of the inputs is debatable (e.g., it is difficult to assign a label at the digit in Figure 1.c).

Since input validity is ultimately a human-related concept, the idea is to model human validity for a certain domain and use such model to guide the generation of valid inputs. Figure 2 presents a possible application of the model by integrating it into a typical reinforcement learning pipeline (i.e., reward model).

Tasks

For the realization of this project, the candidate is expected to perform a series of tasks, among which: (1) familiarize with the task of handwritten digit classification using supervised learning; (2) familiarize with existing test input generators for DL; (3) design an experiment to train a reward model using human-labeled data and possibly integrate the trained reward model into an existing test input generator.

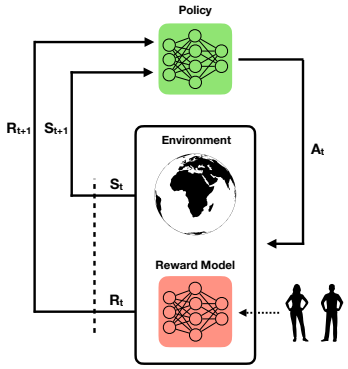


Figure 2: Overview of the approach.

Prerequisites

We are looking for a student who is passionate about DL and motivated to contribute to the project. Knowledge about DL is not required, although it is appreciated. We highly value the willingness to learn these technologies and adapt them depending on the project's needs. During the project, the candidate can rely upon an existing software infrastructure that integrates existing test input generators for different classification tasks (including digit classification). We will work closely with you and provide assistance when needed.

Why You Should Choose This Project

This project is unique in its multi-disciplinary nature: you will learn and use concepts in machine learning (ML), especially DL, software engineering and the ability to set up an experimental setting to rigorously compare different methods.

Upon successful completion of the project, you will have contributed to an open-source project that will allow developers to test their DL models using valid and hence reliable inputs. On your CV side, you will be able to show practical experience with training and above all evaluating DL models.

Further Information

The proposed work is part of the Precrime (Self-assessment Oracles for Anticipatory Testing) ERC project. The interested student can find more information about Precrime on the project's website: https://www.pre-crime.eu/. Interested in this project? Any still-unanswered questions? Drop us an email now!

Test input prioritisation for DL systems

Paolo Tonella, Tahereh Zohdinasab
TAU research group, Software Institute

Abstract

Deep Learning (DL) systems have achieved unprecedented success in solving complex tasks, including safety critical ones such as self-driving. It is crucial to guarantee the reliability of DL systems by testing how they behave with different inputs. Testing DL systems is challenging due to their huge input space, e.g., all possible road configurations for a self-driving car. Moreover, testing complex DL systems may need the execution of expensive simulations. Consequently, prioritising test inputs by choosing only 'high quality' ones, i.e., tests more likely to trigger a failure, is necessary for test cost reduction.

Multiple test prioritisation techniques have been proposed by Software Engineering researchers. However, none of them considered the uniqueness of prioritised test inputs, which guarantees the evaluation of the DL system under different circumstances.

Feature maps (i.e., maps where the inputs are positioned based on their characteristics) can report useful details about a test set, such as the features corresponding to tests that triggered misbehaviours or the probability of observing a misbehavior for each feature combination. In this work, we propose to use feature maps to prioritise diverse, critical test inputs.

The goal of this project is to propose a new approach for test prioritisation for DL systems based on feature maps. The first task of the candidate is to design a test prioritisation strategy that leverages feature maps. The second task is to replicate state of the art approaches for test prioritisation for DL systems and compare them to the proposed approach.

Contacts: paolo.tonella@usi.ch, tahereh.zohdinasab@usi.ch

In-depth Performance Analysis of the Java Vector API

The Java Vector API is an incubator module introduced in JDK 16, allowing developers to express vector computations and eventually compile them to vector hardware instructions at runtime. In addition to a clear and concise API, the module promises high performance and portability, potentially making it a valid alternative to writing vectorized native code.

The performance of the Java Vector API has been analyzed recently by our research group [1]. In particular, the work introduces the first benchmark suite for the Java Vector API (JVBench) and uses it to analyze different aspects of the API:

- Execution time and portability of the API on multiple architectures supporting different vector instruction sets.
- Performance comparison of code generated by the API w.r.t. scalar code as well as code auto-vectorized by the Just in Time (JIT) compiler.
- Patterns and anti-patterns on the use of the API significantly affecting application performance.

The goal of this project is to extend the analyses conducted in our prior work, focusing on aspects of the Java Vector API that have not been studied yet. Activities of interests include:

- Expand the coverage of JVBench by designing and adding new benchmarks that exercise ByteVector and ShortVector, Shift/Rotate operations, and shape-changing operations.
- Evaluate JVBench on ARM AArch64 architectures.
- Evaluate the memory overhead of the Java Vector API.
- Extend the current evaluation of JVBench on different JVM implementations, such as OpenJ9 and GraalVM.
- Evaluate the performance of JVBench against a semantically equivalent C implementation.
- Profile JVBench to obtain the number of executed vector instructions (vs. scalar instructions).

The project is a unique opportunity for the student to deepen the knowledge in the domains of dynamic program analysis, vector instructions, JVM architecture, and empirical evaluation, all being important skills for a software engineer. The student will work side-by-side with the members of the Dynamic Analysis Group at USI, and will receive support in learning advanced topics that will strengthen the abilities as a computer scientist. Applicants interested in this project should be enrolled in the BSc program, have a good knowledge of the JVM architecture, Java, and UNIX-based operating systems, excellent programming skills, and deep interest in the field of dynamic analysis.

Advisors: Prof. Walter Binder and Dr. Andrea Rosà
Assistant: Matteo Basso

References

[1] “Java Vector API: Benchmarking and Performance Analysis”. Basso et al., CC 2023.

Understanding Performance Variability of JVM Workloads

Running the same workload multiple times in separate Java Virtual Machine (JVM) processes often exhibits significant variability in performance metrics. This behavior can generally be explained by nondeterminism, which affects many components involved in the application execution. Among them, dynamic compilation plays a major role in determining application performance. Due to nondeterminism, in multiple runs of the same workload the just-in-time (JIT) compiler may apply different optimization decisions, which in turn may lead to significant differences in execution time.

The goal of this project is to better understand the reasons behind performance variability in different executions of a workload running on the JVM. We aim at investigating in-depth the correlation between an observed execution time and different dynamic metrics, with particular focus on those related to the JIT compiler.

To achieve the goals of the project, the student will be involved in several activities:

1. Identification of workloads where the execution time varies significantly in different runs.
2. Identification of dynamic metrics that can be correlated with the execution time of applications running on the JVM.
3. Development of efficient and accurate profilers to collect metrics of interest.
4. Analysis of the correlation between execution times and metrics of interests in the considered applications.
5. Detection of the root causes of performance variability.

The project is a unique opportunity for the student to deepen the knowledge in the domains of dynamic program analysis, JVM architecture, dynamic compilation, the system stack, and empirical evaluation, all being important skills for a software engineer. The student will work side-by-side with the members of the Dynamic Analysis Group at USI, and will receive support in learning advanced topics that will strengthen the abilities as a computer scientist. Applicants interested in this project should be enrolled in the BSc program, have a good knowledge of the JVM architecture, Java, C/C++, and UNIX-based operating systems, excellent programming skills, and deep interest in the field of dynamic analysis.

Advisors: Prof. Walter Binder and Dr. Andrea Rosà
Assistant: Matteo Basso

Analysis and Optimization of the Code Generated by SparkSQL

Data processing systems such as Apache Spark or Apache Flink are becoming de-facto standards for distributed data processing. One of the key advantages of such systems over their predecessors (e.g., Hadoop) is the availability of high-level programming models such as SQL and the DataFrame API. These interfaces are implemented in SparkSQL relying on query compilation into Java source code through the so-called *whole-stage code generation*.

Unfortunately, debugging the code generated at runtime as well as analyzing its performance is currently a tedious process. This project aims at solving this issue. In a first step, the student involved in this project will study the whole-stage code-generation implementation in Spark and Janino, the Java compiler used by Spark for compiling Java source code into bytecode at runtime. Then, the student will be involved in the following activities:

1) Developing a tool that makes it practical to analyze the code generated by Spark. The tool should dump the code generated by Spark and compiled with Janino for a certain workload such that it can be inspected, analyzed, and customized (either manually or automatically). The tool should also interface with Spark to detect the re-execution of the same workload and use the customized source files instead of generating new ones.
2) Designing such a dump and re-run mechanism in a way which is as non-intrusive as possible, ideally with a Spark plugin, i.e., without modifying the Spark code base.
3) Integrating the implemented tool in an IDE, e.g., as a plugin for Eclipse or IntelliJ IDEA.
4) Using the tool to analyze the generated code for SQL workloads (e.g., TPC-H) to find performance bottlenecks and missing optimization opportunities.

This project is an opportunity for the student to practice with the widely used framework Apache Spark and its code-generation module implemented in SparkSQL, i.e., the query compiler and Janino. The outcome of this project is a tool that makes it easier for Spark developers and researchers to debug and analyze the performance of the code generated by Spark. Moreover, the project will involve a preliminary performance analysis on SQL workloads on dynamic metrics such as e.g. execution time and memory consumption.

Advisors: Prof. Walter Binder, Dr. Filippo Schiavio

Developing an AR/VR application for creating more immersive museum experiences for people with intellectual disabilities

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

This summer project aims to create cutting-edge augmented reality (AR) and virtual reality (VR) technology specifically designed for people with intellectual disabilities. This project will bring together experts in AR/VR technology, people with intellectual disabilities, museum, and educational professionals to co-create innovative solutions that enhance access and inclusion in learning and cultural spaces.

The project team will evaluate the impact of the AR/VR technology on the learning outcomes, accessibility, and engagement of individuals with intellectual disabilities and make necessary adjustments to improve the technology. The goal of this project is to demonstrate the potential of AR/VR technology to create a more inclusive and accessible world for people with intellectual disabilities.

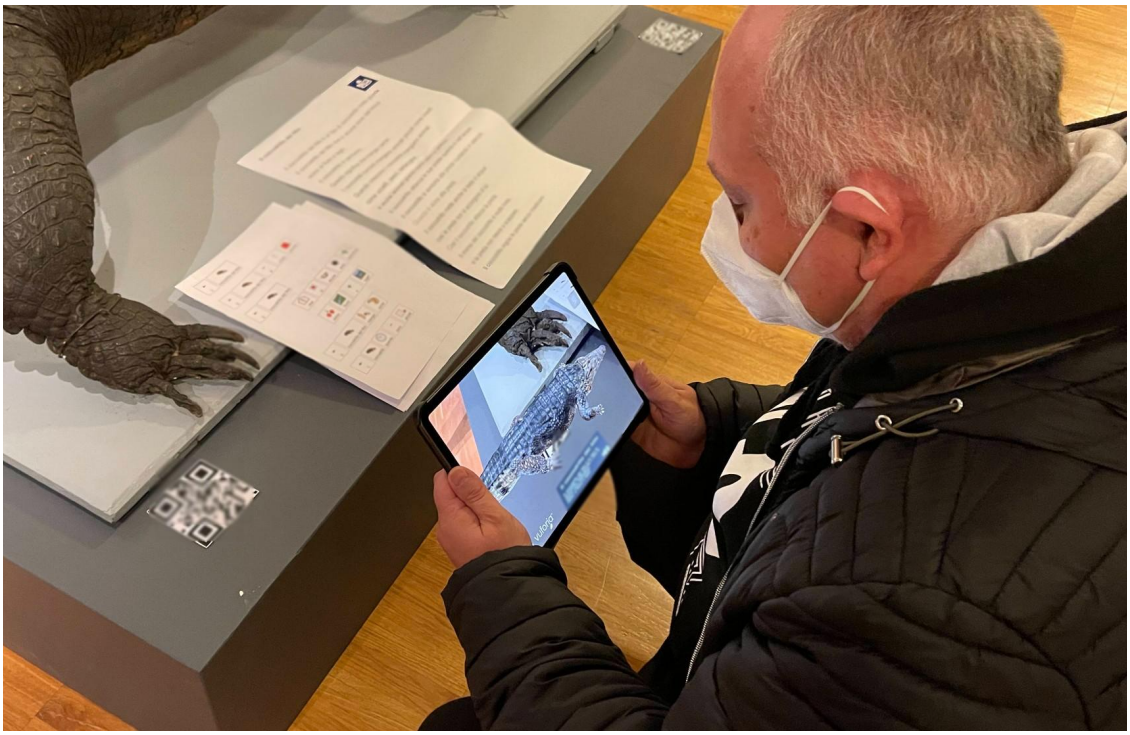


Image: Participant interacting with a 3D model in a museum visit.

You will contribute by supporting the current project, designing, and developing solutions. This will aid users in understanding museum content and enhancing their enjoyment of technology.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. The work can involve researchers from different research areas and countries (Italy and Australia).

Are you interested in this project? Contact us if you have any questions, we will be happy to answer them.

Designing an accessible game for people with intellectual disabilities

UROP project proposal at the Università della Svizzera italiana (USI), Lugano, Switzerland

This summer project focuses on developing accessible games for people with intellectual disabilities. The project team is composed by accessibility experts, educators and a psychologist. The goal is to create an accessible game that is not only enjoyable, but also promotes independence.

The team will work together to provide the specific accessibility needs of people with intellectual disabilities and integrate these into the game design. The game will further require testing and evaluation to ensure that it is simple to use and understand. This summer project hopes to create a more inclusive and accessible world for people with intellectual disabilities.

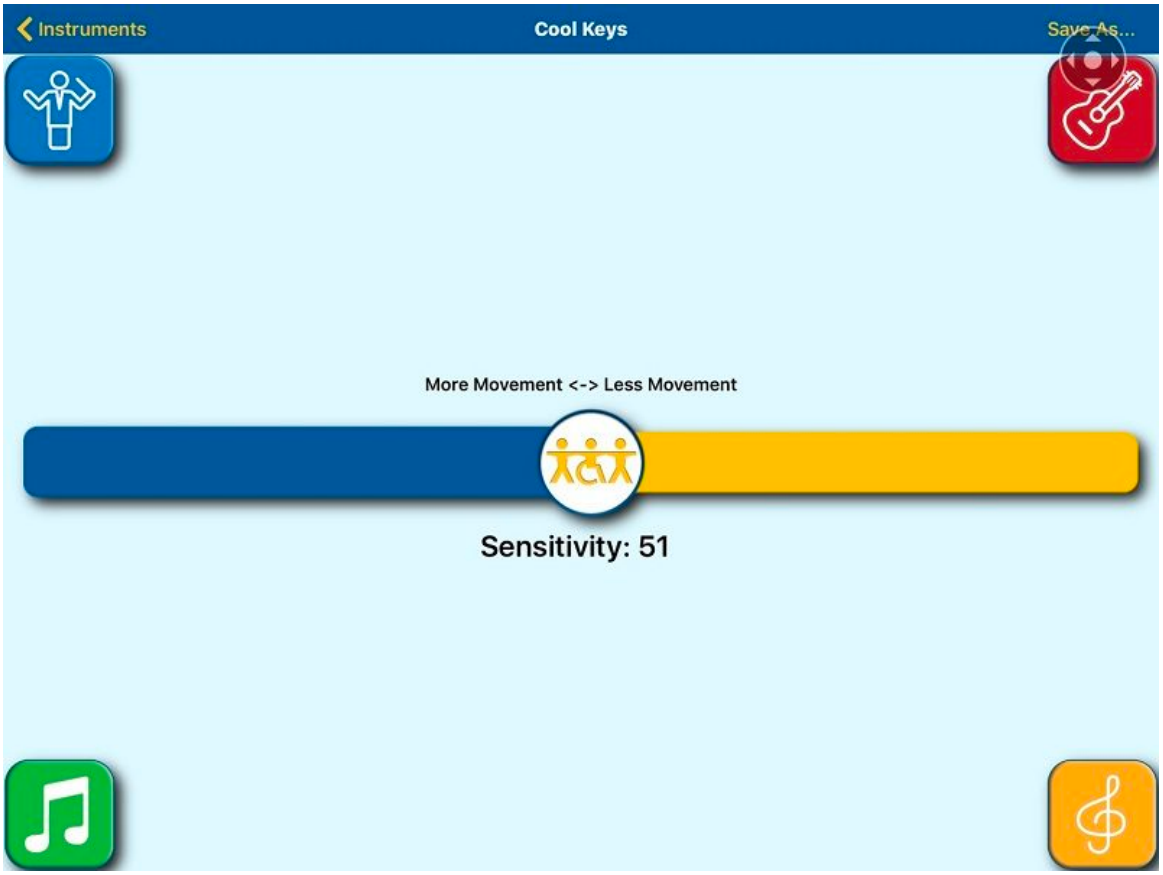


Image courtesy: Waveband - Disability Horizons

You will contribute by supporting our ongoing project, designing, and developing a new fun and accessible game. This will aid users in using innovative and accessible technology.

You will be under the supervision of Dr. Monica Landoni and Leandro Guedes, a Doctoral assistant at USI. The work can involve researchers from different research areas and countries (Italy and Australia).

Are you interested in this project? Contact us if you have any questions, we will be happy to answer them.

Can computational economics predict resilient pathways in the presence of shocks?

Contact Prof. Michael Multerer (michael.multerer@usi.ch), Davide Baroli (davide.baroli@usi.ch)

Due to the introduction of renewable energy sources and other actions in energy transition, global energy markets are subject to huge uncertainty. To accurately predict such markets, corresponding mathematical models must accurately represent the underlying financial and economic phenomena as well as the uncertainty [3]. The objective of this project is to formulate a robust modelling framework [1,2,4] and to define stochastic pathways by employing stochastic differential equations (SDEs). SDEs are crucial to model prices of energy, commodities, and to develop policies for regulators or financial products and strategies [3,4].

Within this project, we are concretely interested in developing an SDEs model for energy markets considering uncertainty in price, demands, availability of energy resources, prosumers action, decentralization of production, shocks (supply chain, conflicts, demographic change, environment degradation and climate change) and the prediction of possible strategies for energy traders. In a follow up step, this model will be integrated into an energy game that is developed with the SFOE SWEET project “Sustainable and Resilient Energy for Switzerland” (SURE, https://sweet-sure.ch/).

The numerical realization of this project will be performed with the help of the ICS cluster (on GPUs for running experiments). In view of the integration into the SURE framework, this project can be that starting point fora master thesis and possibly a scientific presentation at a workshop or a publication in peer reviewed journal.

Prerequisites:

- Knowledge of programming (Python)

- Knowledge of computational methods for solving PDEs and SDEs

- Knowledge of Finance modelling

References:

1. Carmona, Rene, and Xinshuo Yang. Joint Stochastic Model for Electric Load, Solar and Wind Power at Asset Level and Monte Carlo Scenario Generation" arXiv preprint arXiv:2209.13497 (2022).
2. Carmona, René, Gökçe Dayanıklı, and Mathieu Laurière. "Mean field models to regulate carbon emissions in electricity production." Dynamic Games and Applications 12.3 (2022): 897-928.
3. Aïd, René. Electricity derivatives. Cham: Springer International Publishing, 2015.
4. Carmona, René. Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications. Society for Industrial and Applied Mathematics, 2016.
5. Aïd, R., Campi, L., Langrené, N.: A structural risk-neutral model for pricing and hedging power derivatives. Math. Finance. 2022.

GAN for solving fluid flow in aerodynamics problems

Contact: Prof. Michael Multerer (michael.multer@usi.ch), Wei Huang (wei.huang@usi.ch), Davide Baroli (davide.baroli@usi.ch)

In the last decade, the interest in data-centric methods [7], bridging deep neural network, inference from data, partial differential equations (which have been milestone for a wide range of science and engineering models) has drastically increased. Applications range from blood flow simulation over aerodynamics to climate forecasting.

As computational scientists active in the domain of uncertainty quantification for physical problems, we are interested in predicting flow-based simulations [1] in the presence of uncertainty in the input parameters, like boundary conditions or computational shape which is in practice caused by missing data or the data acquisition procedure, for example low quality images.

In recent years, GAN-based Physics informed neural networks [4,6] and reduced order methods [2,3] have shown the potential to address fluid-based PDEs with uncertain input. During this summer project, you will investigate GAN architectures and their corresponding latent spaces, to infer posterior probability distributions from data and solve inverse problems [5].

Prerequisites:

- Knowledge of programming, PyTorch (preferable) or other AI.

- Knowledge of Aerodynamics (fluid-based PDEs), not necessary but highly appreciated

- Knowledge of computational methods for solving PDEs and SDEs

References:

[1] Berselli, L.C., Iliescu, T., & Layton, W.J. (2005). Mathematics of Large Eddy Simulation of Turbulent Flows.

[2] Koc, Birgul, Changhong Mou, Honghu Liu, Zhu Wang, Gianluigi Rozza and Traian Iliescu. “Verifiability of the Data-Driven Variational Multiscale Reduced Order Model.” *Journal of Scientific Computing* 93 (2021)

[3] Xie, X., Webster, C., & Iliescu, T. (2019). Closure Learning for Nonlinear Model Reduction Using Deep Residual Neural Network. *Fluids*.

[4] Xie, You, Erik Franz, Mengyu Chu and Nils Thürey. “tempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow.” *ACM Trans. Graph.* 37 (2018): 95.

[5] Ghosh, P., Zietlow, D., Black, M.J., Davis, L., & Hu, X. (2021). InvGAN: Invertible GANs. *German Conference on Pattern Recognition*.

[6] Kim, Byungsoo, Vinicius C. Azevedo, Nils Thürey, Theodore Kim, Markus H. Gross and Barbara Solenthaler. “Deep Fluids: A Generative Network for Parameterized Fluid Simulations.” *Computer Graphics Forum* 38 (2018).

[7] Shukla, K., Xu, M., Trask, N., & Karniadakis, G. E. (2022). Scalable algorithms for physics-informed neural and graph networks. *Data-Centric Engineering*, *3*, e24.

Three-dimensional implementation of a cell-by-cell model for cardiac electrophysiology

The cardiac tissue is composed, among others, of cardiomyocyte cells, in which electric potential propagates and generates muscle contraction. In healthy hearts, tight cellular connections ensure smooth potential propagation and correct heart functioning. In cardiac electrophysiology, the cell-by-cell models describe this phenomenon, i.e. the action potential's propagation through the cardiac tissue at the cellular level. Therefore, these models are particularly important to study heart diseases, as fatty tissue infiltration in between cells.

The goal of this project is to implement a cell-by-cell model in three dimensions using a boundary integral approach, and therefore the boundary element method (BEM). The BEM allows to represent the problem only at the cellular surfaces. Therefore, it requires significantly fewer degrees of freedom than the popular finite element method, which would solve the problem in the interior of the cells as well.

For the two-dimensional problem, a prototypical BEM code developed at the Euler Institute already exists. This code implements the basic BEM operators for single cells and the routines needed for cell-to-cell communication. Starting from existing three-dimensional codes for the BEM, which already implement the basic operators, the student would implement the missing cell-to-cell communication routines.

This project has an important practical impact since it paths the way for realistic cell-by-cell model simulations. Nowadays, three-dimensional cell-by-cell models are simulated using the finite element method and are overly expensive for practical applications, we expect that a BEM solution would be significantly cheaper.

The successful candidate can grow scientifically and extend the project outcomes to a master thesis. Furthermore, the student will be able to benefit from existing numerical codes and computational resources at the Euler Institute.

Prerequisites
Understanding of C++ programming language and linear algebra. A basic understanding of scientific computing and calculus (function, derivatives) is welcome.

Contact
Giacomo Rosilho de Souza giacomo.rosilhodesouza@usi.ch, Simone Pezzuto simone.pezzuto@usi.ch, Rolf Krause rolf.krause@usi.ch. Euler Institute.

Setting a realistic simulation environment for testing self-driving cars

Paolo Tonella, Nargiz Humbatova
Software Institute@USI
paolo.tonella@usi.ch,nargiz.humbatova@usi.ch

In the last decade, Deep Learning (DL) solutions are adopted in a constantly growing number of domains. DL applications influence important aspects of life by tackling tasks such as fraud detection or face recognition, while some are employed in safety critical domains such as autonomous driving. This makes reliability and robustness of such systems of a high importance. While testing such systems in the real-world environment ensures the reliance of the results, it becomes increasingly expensive for complex systems. For example, providing an autonomous vehicle with all kinds of possible road situations to evaluate its safety is non-trivial and sometimes even an impossible task. Simulation testing is an efficient way to test systems like autonomous vehicles before conducting resource-demanding real-word trials.

CARLA

CARLA is an open-source autonomous driving simulator [1]. It was developed specifically to serve the needs of researches in development, training, and testing of various autonomous driving systems. CARLA comes with the flexibility in customisation of the tool and wide range of assets (urban layouts, buildings, vehicles, pedestrians, street signs, etc.) that aid generation of various testing scenarios. It is based OpenDRIVE standards to define the urban setting and Unreal Engine to run the simulation. Moreover, CARLA allows flexible setup of sensors typical to autonomous vehicles (such as cameras and LIDARs) and provides users with features such as usage of GPS coordinates, accelerations to enable training of different driving strategies. On Fig. 1 is provided an example of the same simulation scenario in CARLA simulator under 4 different weather conditions.

Project Proposal

The goal of the project is to set up and prepare a simulation environment in CARLA for testing autonomous cars. The environment should include pedestrians, vechicles, traffic signs, and other reasonable obstacles. It also involves adopting a Deep Learning system emulating an autonomous car that is able to successfully perform in the described environment. The last step would be to integrate the calculation of various driving quality metrics in the simulator. The examples of such metrics could be speed, lane position, acceleration, and etc. [3]. This project will prepare grounds for further studies in the domain and academic contributions.

In the frame of this project, the student will learn about state-of-the-art simulation practices in the domain of autonomous vehicle testing, will practice with popular DL frameworks and model architectures suitable for the task of autonomous driving. Moreover, the student will be able to have practice in data collection and training of a DL model and later employing the trained model in the simulator.

Figure 1: A street from CARLA simulator from a third-person view in 4 different weather conditions (clear day, daytime rain, daytime shortly after rain, and clear sunset)[2]

Additional Information

The project will be carried out within the TAU research group at the Software Institute (https://www.si.usi.ch) and contribute to the PRECRIME ERC research project (https://www.pre-crime.eu). Students are supervised by researchers of the TAU group who follow them constantly and provide them with timely feedback, advice and directions. The code developed for the projects is typically released as an open source project and the results are often included in scientific publications. Both code and publication would contribute to a stronger CV of the participating student.

REFERENCES

[1] 2023. CARLA. https://carla.org/
[2] Alexey Dosovitskiy, Germán Ros, Felipe Codevilla, Antonio M. López, and Vladlen Koltun. 2017. CARLA: An Open Urban Driving Simulator. *ArXiv* abs/1711.03938 (2017).
[3] Gunel Jahangirova, Andrea Stocco, and Paolo Tonella. 2021. Quality Metrics and Oracles for Autonomous Vehicles Testing. In *2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST)*. 194–204. https://doi.org/10.1109/ICST49551.2021.00030

Towards an Intelligent Post-training Mutation Tool for Deep Learning Systems

Paolo Tonella, Nargiz Humbatova
Software Institute@USI
paolo.tonella@usi.ch,nargiz.humbatova@usi.ch

Deep Learning (DL) has become an integral part of many groundbreaking projects and products we use everyday. As quality and safety remains the main concern for the developers and users of modern products based on Artificial Intelligence, different techniques aimed at assessing their quality are of increasing interest for research community.

Mutation Testing

Mutation testing is a technique that deliberately seeds faults in form of small syntactic changes into the program under test to create a set of faulty programs called mutants. The general principle underlying this approach is the assumption that faults used by mutation testing represent the mistakes that programmers usually make. Mutation testing aims to assess the quality of a given test suite in terms of its capability to detect faults. For this, the test suite is executed on each of the generated mutants. If the result for a given mutant is different from the result of running the original program then the mutation is considered killed. The ratio of killed mutants to the overall number of generated mutants is called mutation score. The higher the mutation score, the better is the quality of the test suite.

The example in Figure 1 shows a method `subtract` that subtracts two integer values and returns the result. It has two mutations: in `Mutant 1` the subtraction is replaced with multiplication and in `Mutant 2` it is replaced with addition. If our test suite has only `test0`, none of the two mutations would be killed (as they all return the expected value `0`) and the mutation score is 0%. If we add test case `test1()` to our test suite, then `Mutant 1` gets killed and the mutation score becomes 50%. Once we add test case `test2()`, both mutations get killed and the mutation score achieves its maximum value of 100%.

Mutation Testing for DL Systems

In traditional software systems the decision logic is often implemented by software developers in the form of source code. In contrast, the behaviour of a DL system is mostly determined by the training data set and the training program, i.e. these are the two major sources of defects for DL systems. Thus, there should be a specific approach to mutation testing of DL systems. There currently exist two tools that are designed specifically for performing mutation testing for DL systems. However, one of the tools is a pre-training one, which means it injects the faults into system prior to the training and thus is computationally expensive, while the second one, a post-training mutation tool, injects faults that are random and not very likely to happen in real world. Such faults usually introduce slight noise or modifications to a randomly selected subset of weights or change a structure of an already trained DL model by adding/deleting its layers or replacing the activation function.

```
// Original Program
public int subtract(int a, int b) {
   return a - b;
}
```

```
// Mutant 1
public int subtract(int a, int b) {
   return a * b;
}
```

```
// Mutant 2
public int subtract(int a, int b) {
   return a + b;
}
```

```

public void test0() {
   assertEquals(0, subtract(0, 0));
}

public void test1() {
   assertEquals(-4, subtract(-2, 2));
}

public void test2() {
   assertEquals(1, subtract(2,1));
}
```

Figure 1: Mutation Testing Example

Currently, mutation testing is being applied to various tasks for DL systems such as program repair, generation of optimal oracles for self-driving cars, detection of adversarial inputs, prioritisation of test inputs for the labelling, etc. Availability of a mutation tool that can inject changes that resemble the effect inflicted by real faults would be extremely useful also for these approaches as well as the advance of DL testing in general.

Project Proposal

The goal of the project is to develop a new post-training mutation tool that would solve the limitations set by the previous approaches, i.e. to introduce smarter and fast DL-specific mutation operators that produce stable and reliable results and would facilitate the increased interest for mutation testing in DL community.

In the frame of this project, the student will learn about state-of-the-art techniques in the domain of mutation testing for DL systems, their limitations and advantages. The student will practice with most popular DL frameworks and widely-used models and datasets.

Additional Information

The project will be carried out within the TAU research group at the Software Institute (https://www.si.usi.ch) and contribute to the PRECRIME ERC research project (https://www.pre-crime.eu). Students are supervised by researchers of the TAU group who follow them constantly and provide them with timely feedback, advice and directions. The code developed for the projects is typically released as an open source project and the results are often included in scientific publications. Both code and publication would contribute to a stronger CV of the participating student.

Title: Learning microstructural conduction in the heart

The heart is composed by billions of cells. A numerical simulation of cardiac electric activity at subcellular resolution is therefore out of reach now and, even if it were not, it would not be feasible on a clinical time scale. But we can simulate a small chunk of tissue, a fraction of millimeter cube, and obtain an estimate of the macroscopic properties of the tissue. Such mean-field parameters can then be used in an organ-scale simulation at a feasible mesh resolution. This technique is called homogenization and it is commonly used in computational mechanics, amongst other fields.

The objective of this project is to learn the conductivity tensor of the cardiac tissue from microscale simulation data. The conductivity will depend on the local fiber direction of the tissue, as well as other parameters of the microscale model, for instance gap junction distribution, aspect ratio of myocyte, and the presence of fibrosis. Given the large set of parameters, neural networks could offer an effective approach for encoding the conductivity.

The successful candidate can grow scientifically and to extend the project outcomes to a master thesis. Furthermore, the student will be able to benefit from existing numerical codes and computational resources at the Euler Institute and at the CSCS.

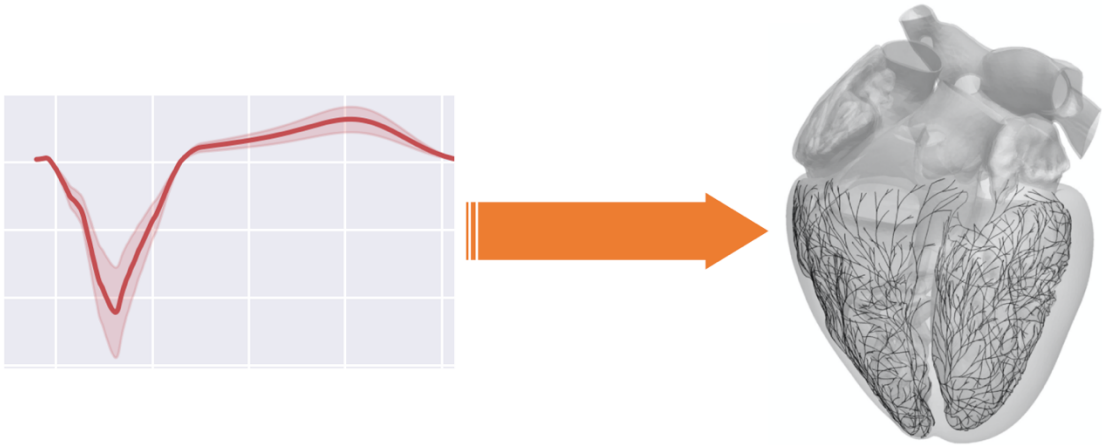
Prerequisites
Basic understanding of understanding of Calculus (function, derivatives) and differential equation. Some experience with Python and C++ is welcome.

Contact
Simone Pezzuto simone.pezzuto@usi.ch, Giacomo Rosilho de Souza giacomo.rosilhodesouza@usi.ch, and Rolf Krause rolf.krause@usi.ch, Euler Institute.

Title: Identification of the Purkinje network via machine learning

The Purkinje network is a tree-like structure present in our heart. The Purkinje network is like a highway system for the electric stimulus that should quickly and synchronously activate the ventricles of the heart. A malfunction of the system, called cardiac electric dysfunction, is a serious cardiac disease affecting millions of people worldwide. It is typically fixed by a pacemaker.

The objective of this project is the identification of this network from a standard 12-lead electrocardiogram (ECG). In practice, we have developed an algorithm to automatically generate physiological Purkinje networks and subsequently simulate the corresponding ECG. Thus, given a patient ECG, we would like to identify the set of parameters of the network that best explains the observed ECG. This is achieved via surrogate modeling of the loss function with Gaussian Process Regression.



The successful candidate can grow scientifically and to extend the project outcomes to a master thesis. Furthermore, the student will be able to benefit from existing numerical codes and computational resources at the Euler Institute and at the CSCS.

Prerequisites
Some experience with Python and machine learning tools. The model works already on Google Colab.

Contact
Simone Pezzuto simone.pezzuto@usi.ch, Rolf Krause rolf.krause@usi.ch, Euler Institute.

Title: Shape2PDE, a Parametric PDE Solver from Shapes

Partial Differential Equations (PDEs) are ubiquitous in mathematical modeling and applications. PDEs like the Navier-Stokes equations can be very expensive to solve: can we do better? A PDE may depend on several parameters including, mostly notably, the domain of the problem. Computationally speaking, a shape can be very general and hard to parametrize. A powerful and compact approach consists in encoding the shape through a signed distance function. For instance, the function $f(x, y, z) = x^2 + y^2 + z^2 - 1$ encodes the unit sphere.

The objective of this project is to learn the mapping from the shape of the domain to the solution of the PDE: succinctly, Shape2PDE. The idea is to combine DeepSDF [1] or Get3D [2], a deep neural network that can encode signed distance functions, with a mesh-free PDE solver. The network has two inputs, a code and a point coordinate and should return the solution of the PDE at the point. The training of the network consists in learning from a large set of shapes, all encoded as signed distance functions, and PDE solutions. DeepSDF is so powerful that can recover a full shape even from a sparse set of points partially covering it.

[FIG]

The successful candidate can grow scientifically and to extend the project outcomes to a master thesis. Furthermore, the student will be able to benefit from existing numerical codes and computational resources at the Euler Institute and at the CSCS.

Prerequisites
Basic understanding of neural networks and Python programming language. Previous experience with PyTorch, TensorFlow or similar is good but not strictly necessary. Basic understanding of Calculus (function, derivatives) is welcome.

References
[1] Park et al., “DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation”, 2019, arXiv: 1901.05103v1
[2] Gao et al., “GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images”, NeurIPS 2022, https://nv-tlabs.github.io/GET3D/

Contact
Simone Pezzuto simone.pezzuto@usi.ch, Hardik Kothari hardik.kotari@usi.ch, and Rolf Krause rolf.krause@usi.ch, Euler Institute.

Self-monitoring using wearable devices

UROP project proposal at the Università della Svizzera Italiana (USI), Lugano. Switzerland

Background

Wearable devices such as -- e.g., smartwatches, smart earbuds, or digital rings -- enable the continuous and unobtrusive collection of sensor data about, e.g., physical activity or heart rate of the persons wearing the devices. This data, in turn, allows to infer the context and behavior of the users and offer them novel applications and services.

A core research activity of the Mobile&Wearable Computing group is the design and development of novel systems for self-tracking using wearable devices with the goal of helping people improve their overall well-being and productivity. In this context, the group often runs in-the-wild studies to collect datasets to analyze and build models of context- and human behavior recognition. To this end, several instruments, including applications for mobile and wearable devices are developed and used.

The aim of this project is to extend an existing mobile application to use in the context of such data collection studies. The app is implemented in Flutter, a cross-platform application development programmed using the Dart language. The foreseen extensions to the app include but are not limited to: (1) including mechanisms to collect data from novel wearable devices (e.g., earbuds, smart rings); (2) embedding machine learning model to recognize basic human activities using sensor data; (3) adding visualization mechanisms to provide feedback to the user (i.e., plots of collected sensor data, notifications, bio-feedback).

Required skills and knowledge

- Good knowledge – or willingness to learn – mobile application programming (e.g., using Android or a cross-platform programming framework like, e.g., Flutter).
- Good knowledge of technologies for data visualization and feedback.
- Basic knowledge of data analysis and machine learning techniques (preferably using Python).

Expected outcomes

- Extend an existing Flutter application with additional mechanisms to collect data from wearable devices.
- Embed machine learning models into the application to automatically recognize basic human behaviors.
- Add adequate data visualization mechanisms to the app.

Supervisors and contact information: Prof. Dr. Silvia Santini, silvia.santini@usi.ch.

A situated self-report device for data collection studies

UROP project proposal at the Università della Svizzera Italiana (USI), Lugano. Switzerland

Background

Wearable devices such as – e.g., smartwatches, smart earbuds, or digital rings – enable the continuous and unobtrusive collection of sensor data about, e.g., physical activity or heart rate of the persons wearing the devices. This data, in turn, allows to infer the context and behavior of the users and offer them novel applications and services.

A core research activity of the Mobile&Wearable Computing group is the design and development of novel systems for self-tracking using wearable devices with the goal of helping people improve their overall well-being and productivity. In this context, the group often runs in-the-wild studies to collect datasets to analyze and build models of context- and human behavior recognition. To this end, several instruments, including so-called situated self-report devices (SSRs), are designed and developed.

An SSR is a small and simple device used to allow participants to data collection studies to provide self-reports about, e.g., their work activities, emotions, or sleep quality – in a quick and uncomplicated manner. An SSR is designed for a specific use and provides few, but very easy- and quick-to use functionalities. A typical SSR may include, e.g., a small screen to visualize a few questions the participants has to answer and a knob to select answers to the questions. SSRs are small objects that can be kept on a participants desk or night table, or even be attached to a door or wall. Several studies have shown that the use of SSRs can increase significantly the number of self-reports provided by participants during data collection studies. This is important since self-reports are used as labels to train machine learning models. The more and the more accurate the labels are, the better the models.

The aim of this project is to design and develop an SSR device for a large data collection study planned in the autumn/winter 2023/2024. An existing SSR used in a previous study can be used as starting point. The student will be required to design and develop an initial prototype of the device, do some preliminary testing, and then develop a final version to be used in the study.

Required skills and knowledge

- Good knowledge – or willingness to learn – low-level programming (e.g., C or C++ to program Arduino boards or simple microcontrollers).
- Some knowledge of prototyping or maker lab experience.

Expected outcomes

- A situated-self-report (SSR) device to be used in the mentioned data collection study.
- Extensive documentation of the design and development of the SSR.
- Documentation about the outcomes of the testing of the SSR.

Supervisors and contact information: Prof. Dr. Silvia Santini, silvia.santini@usi.ch.

DIVERSIM: A SIMULATION TOOL FOR SPECIES EVOLUTION.

Prof. Ernst C. Wit (wite@usi.ch) and Dr. Francisco Richter (richtf@usi.ch)

DESCIPTION

This project is centered around the use of statistical learning algorithms and cutting-edge bioinformatics to better understand the drivers of species evolution. The goal is to develop software that will allow for a more comprehensive study of the factors that contribute to the evolution of species and their interactions with each other and their environment.

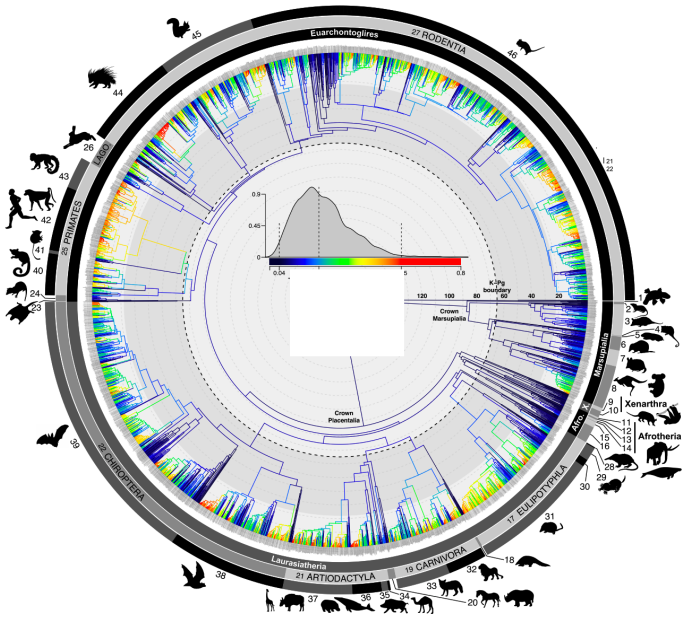


Figure 1: Evolutionary process are usually described by phylogenetic trees. A phylogenetic tree has two components, the topology or shape of the tree, and the time at its nodes. Phylogeny of Mammals provided by Upham et. al. in 2019.

The project will be divided into several stages. Firstly, a brief study of species diversification models and simulation techniques will be conducted.

Next, multiple simulation routines will be coded. Different hypotheses about the drivers of species evolution will serve as input parameters for the simulations. Each hypothesis will then be modeled as a neuron in a large neural network.

The simulation results will be analyzed and used to develop a visualization app that will allow users to study the evolution of species from different points of view. The app will provide an interactive platform for users to compare and analyze simulation results.

The Species Evolution Simulator project is expected to make significant contributions to the field of bioinformatics. The software developed in the project will be a valuable tool for researchers and educators to study the drivers of species evolution. The simulation methodologies and the visualization app will be reported in a technical paper, providing a comprehensive summary of the neural network created and the software capabilities. The project will also generate a database of simulation results that can be used for future research, contributing to a deeper understanding of the drivers of species evolution.