



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--|---|-------------------------|
| Digital Twin for Edge Data Stream Processing | <p>A Digital Twin is a virtual model of physical systems, assets, or environments, continuously updated by real-time data from the physical system to analyze and optimize its behavior. This project aims to support Digital Twinning in edge computing systems by establishing a "physical" system using multiple Raspberry Pis for edge data stream processing.</p> <p>The tasks include:</p> <ol style="list-style-type: none"> 1. Setting up a distributed system using Raspberry Pis. 2. Deploying a state-of-the-art data stream processing engine (e.g., eKuiper, Apache Flink, Apache Storm, or RisingWave) on the Raspberry Pis. 3. Programming and deploying state-of-the-art application benchmarks like RIOTBench. 4. (Optional) integrating with container orchestration technologies such as K3s and KubeEdge. | Georgios Theodoropoulos |
| Muti-Agent Learning in Edge AI | <p>This project aims to investigate multi-agent learning for decentralized control in edge computing. The objective is to implement state-of-the-art algorithms for adaptive edge stream processing to support auto-scaling and dynamic placement of processing operators.</p> <p>The tasks include:</p> <ol style="list-style-type: none"> 1. Modeling the edge stream processing application using state-of-the-art simulators such as iFogSim (https://github.com/Cloudslab/iFogSim) or CloudSim. 2. Implementing state-of-the-art multi-agent learning algorithms for operator scaling and scheduling. <p>Related work: Valeria Cardellini, Francesco Lo Presti, Matteo Nardelli, and Gabriele Russo Russo. 2022. Runtime Adaptation of Data Stream Processing Systems: The State of the Art. ACM Comput. Surv. 54, 11s, Article 237 (January 2022), 36 pages. https://doi.org/10.1145/3514496</p> | |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|------------------------------|---|-------------------------|
| Middleware for Digital Twins | <p>Current technologies for creating Digital Twins (DTs) are limited, with only a few software options available (e.g., Azure Digital Twins, AWS TwinMaker, Siemens MindSphere). These often require premium licenses, restricting their use in research. Open-source alternatives and detailed documentation are needed to support DT development based on current research. Solutions like Eclipse Ditto offer viable alternatives for creating and expanding DTs.</p> <p>Middleware is software that lies between an operating system and the applications running on it. Developing middleware for DTs is challenging due to the variety of DT architectures. A comprehensive, well-documented middleware system is needed to assist researchers. This middleware should enable devices like Raspberry Pi, smart robots, or IoT simulators to create virtual replicas and be reusable for various research purposes.</p> <p>The objective of this project is to develop a middleware to support interaction between various components of a DT-based system, such as data exchange between IoT devices and agent-based simulators, or between multiple DTs. This project will integrate standard technologies such as Eclipse Ditto (https://eclipse.dev/ditto/), message queues (RabbitMQ, etc.), remote procedure call (gRPC, etc.), with configurable APIs. A reference architecture will be provided to facilitate the development.</p> <p>Related work: Andre Almeida, Thais Batista, Everton Cavalcante, Flávia Delicato, Rebeca Motta, and Marco Vieira. 2023. Middleware for Digital Twins: A Systematic Mapping Study. In Proceedings of the 1st International Workshop on Middleware for Digital Twin (Midd4DT '23). ACM, New York, NY, USA, 19–24. https://doi.org/10.1145/3631319.3632302</p> | Georgios Theodoropoulos |



| Title | Introduction | Supervisor |
|--|--|-------------------------|
| Large Language Model for Simulation Interpretation | <p>Agent-based simulation is widely used to analyse complex systems such as smart cities, but they require expert knowledge for modeling and result analysis, making the configuration and output difficult for ordinary users to understand. Large Language Models (LLMs) like GPT-4 and Llama 3 excel at comprehending and generating text-based content. This project aims to leverage LLMs to interpret simulation analysis results in natural language.</p> <p>The project will use Retrieval Augmented Generation (RAG) technology to build an AI agent capable of parsing outputs and parameters from simulators, facilitating Q&A dialogues to help users understand simulation results.</p> <p>Tasks involved:</p> <ol style="list-style-type: none">1. Deploy an LLM (e.g., Llama 3) locally, without additional training.2. Use LangChain or a similar framework to set up the RAG workflow and knowledge base.3. Connect the workflow to a simulator to obtain simulation results.4. Demonstrate a proof-of-concept. <p>Related work:</p> <p>N. Zhang, C. Vergara-Marcillo, G. Diamantopoulos, J. Shen, N. Tziritas, R. Bahsoon, and G. Theodoropoulos, ‘Large Language Models for Explainable Decisions in Dynamic Digital Twins’ , 2024, arXiv preprint, https://arxiv.org/abs/2405.14411</p> <p>N. Zhang, R. Bahsoon, N. Tziritas, and G. Theodoropoulos, ‘Explainable Human-in-the-loop Dynamic Data-Driven Digital Twins’ , in <i>4th International Conference on Dynamic Data Driven Applications Systems (DDDAS2022)</i>, Nov. 2022. https://arxiv.org/abs/2207.09106</p> | Georgios Theodoropoulos |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--|--|-------------------------------|
| Designing Evolutionary Multi-Objective Optimization Algorithms Using Large Language Models | Designing evolutionary multi-objective optimization algorithms (EMOAs) involves selecting and configuring various components such as selection mechanisms and genetic operators. This process is traditionally manual and labor-intensive. Leveraging large language models (LLMs), such as GPT-4, offers an opportunity to automate and enhance the design process, potentially discovering novel configurations that human designers might overlook. This project focuses on designing and implementing a framework that integrates large language models to automate the design of EMOAs. Some basic knowledge of multi-objective optimization, evolutionary multi-objective optimization algorithms, and large language models is required to complete this project. | Hisao Ishibuchi/Lie Meng Pang |
| Enhancing Evolutionary Multi-Objective Optimization Algorithms Using Hyper-Heuristics | The aim of this project is to improve evolutionary multi-objective optimization (EMO) algorithms using hyper-heuristic approaches. The main objective is to identify the key algorithmic components of existing EMO algorithms and investigate their effectiveness in generating high-quality solutions. The project will employ hyper-heuristic strategies to explore different combinations of algorithmic components and utilize an unbounded external archive to store and analyze the quality of the generated solution sets. Additionally, the project will explore the possibility of training models using data from the unbounded external archive to generate high-quality solutions. | Hisao Ishibuchi/Lie Meng Pang |



| Title | Introduction | Supervisor |
|-----------------|--|-------------------------------|
| 基于头脑风暴优化的设备协同学习 | <p>The rapid proliferation of IoT devices and the evolution of digital technologies have brought the research spotlight to the domain of device-server/device-device collaborative learning [1]. This paradigm allows decentralized devices to learn a global model collaboratively, thereby enabling them to make intelligent decisions independently. Nevertheless, efficiency and performance optimization of such learning methods remain a daunting challenge, primarily due to the constraints of resources and computational power in each device. The principal aim of this project is to optimize the device-edge/device-device collaborative learning process by integrating the Brain Storm Optimization Algorithm (BSO) [2]. The three phases of the project are as follows:</p> <ul style="list-style-type: none">• Phase 1 (Learning and Replication): understanding and replicating the BSO algorithm and basic on-device learning algorithms;• Phase 2 (Integration): integrating the BSO algorithm with on-device learning algorithms;• Phase 3 (Optimization): improving the performance of the algorithm implemented in Phase 2 in terms of the speed, accuracy, and resource efficiency. <p>References:</p> <p>[1] Dhar, Sauprik, et al. "A survey of on-device machine learning: An algorithms and learning theory perspective." <i>ACM Transactions on Internet of Things</i> 2.3 (2021): 1-49.</p> <p>[2] Cheng, Shi, et al. "A comprehensive survey of brain storm optimization algorithms." <i>2017 IEEE Congress on Evolutionary Computation (CEC)</i>. IEEE, 2017.</p> | Chair Professor: Yuhui Shi |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|-----------------|---|-------------------|
| 基于推荐系统的头脑风暴优化算法 | <p>本项目旨在通过推荐系统来优化和提高头脑风暴优化算法的性能，分三个阶段开展研究工作：</p> <ul style="list-style-type: none">• 第一步: 推荐系统实现;• 第二步: 头脑风暴优化算法实现；• 第三步：基于推荐系统的头脑风暴优化算法实现。 | 史玉回，讲席教授 |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--------------|--|-------------------|
| 群体机器人动态目标围捕 | <p>本项目旨在通过群体智能算法来实现多机器人围捕，分三个阶段开展研究工作：</p> <ul style="list-style-type: none">• 第一步：群体机器人动态单目标围捕；• 第二步：群体机器人动态多目标围捕；• 第三步：基于头脑风暴优化算法框架的动态多目标围捕。 | 史玉回，讲席教授 |



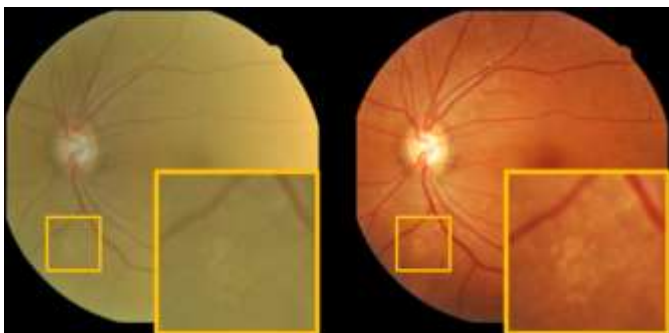
Imed创新实验介绍

2023年9月学期

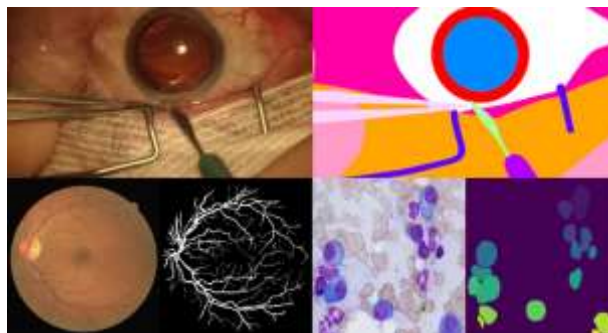
可信辅助诊疗组

带队人：李衡，刘江

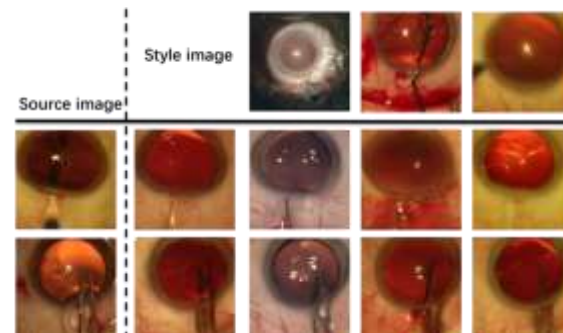
项目内容



图像增强



分割检测



图像生成

题目

1. 基于分布式数据的医学图像分割
2. 白内障患者眼底影像增强系统
3. 基于生成模型的医学数据增广
4. 弱监督医学图像异常检测

迁移学习

领域泛化

对比学习

联邦学习

自监督

知识蒸馏

预期成果

南方科技大学
大学生创新创业训练计划项目
申请书

项目名称：基于网络数据的白内障手术数据分析
项目类别：创新训练
专业名称：计算机科学与技术
指导教师：刘江 职称：教授
申请日期：2022 年 6 月 1 日

项目
大创项目

MVD-Net: Semantic Segmentation of Cataract Surgery Using Multi-View Learning
Abstract: Cataract surgery is a common ophthalmic surgery. The segmentation of cataract surgery images is a challenging task. In this paper, we propose a multi-view learning (MVD) framework for cataract surgery image segmentation. The framework consists of a multi-view learning module and a segmentation module. The multi-view learning module is designed to learn the relationship between different views of the same image. The segmentation module is designed to segment the cataract surgery images. The framework is evaluated on a public dataset and achieves state-of-the-art results.

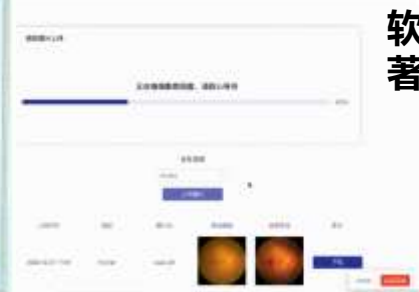
学术论文



发明专利



软件著作

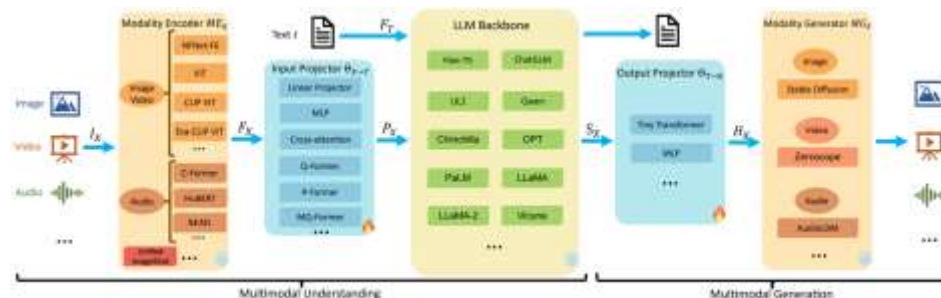


AI+教育组

带队人：聂秋实、魏嘉祺，刘江



学习
报告
生成



大语言模型



评估标准设计

- 可选课题
1. 基于大语言模型的学生报告打分方法研究
 2. 面向人工智能导论课程的自适应学习平台
 3. 南科大计算机系的进系时间评测系统
 4. “人际+人机”混合模式下的学生的学习效果智能评估平台研发

大语言模型

transformer

AI+教育

未来教育

人工智能

“人际+人机”

预期成果



学术论文

| | | | | | |
|----|-----------|--------|---------|----|-------------------|
| 24 | 计算机科学与工程系 | 聂秋实、李科 | 自然科学类 | 课题 | 基于预训练模型的自适应推荐系统 |
| 24 | 计算机科学与工程系 | 魏嘉祺、李科 | 科技发明制作类 | A类 | 融合不确定性信息的智能决策支持系统 |
| 24 | 计算机科学与工程系 | 魏嘉祺、李科 | 科技发明制作类 | 项目 | 基于预训练模型的自适应推荐系统 |

| | | | | | | |
|----|----|----------|-----------|-------------------|-----|-------|
| 24 | 项目 | 智能决策支持系统 | 计算机科学与工程系 | 基于预训练模型的自适应推荐系统开发 | 曹鸣 | 刘江 |
| 24 | 项目 | 智能决策支持系统 | 计算机科学与工程系 | 基于预训练模型的自适应推荐系统 | 魏嘉祺 | 刘江、曹鸣 |

大创攀登

项目

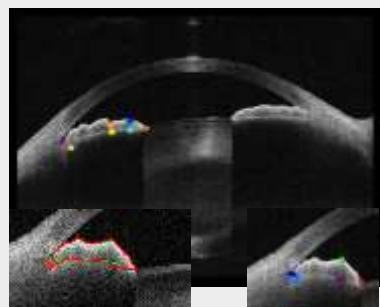


软件著作权

AS-OCT医疗影像分析与低视力辅助导盲

参与人：

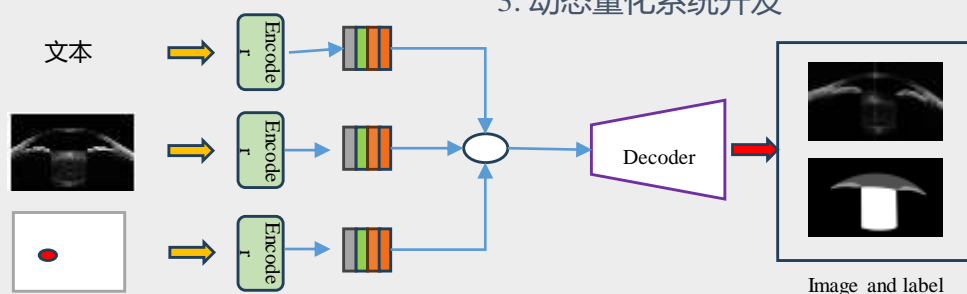
小组：智能图像分析与设备组



基于AS-OCT动态视频的虹膜位置追踪

动态AS-OCT 图像能完整清楚地呈现虹膜受光刺激后的弹性运动，在青光眼开闭诊断中可以提供重要信息。需要开发相应追踪算法。

1. 实时视频分割算法
2. 虹膜位置关键点追踪
3. 动态量化系统开发



基于生成式人工智能技术的模型性能提升

1. 编码文本、图像以及prompt提示；
2. 编码对齐，并基于对齐后的编码生成数据与标签；
3. 迁移大语言模型或视觉模型的通用知识；

基于机器狗的导盲辅助机器人

1. 如何结合图像识别算法到机器狗平台上，如何将感知信息传递给低视力人群？



机器狗 + 图像识别

信息提示方法

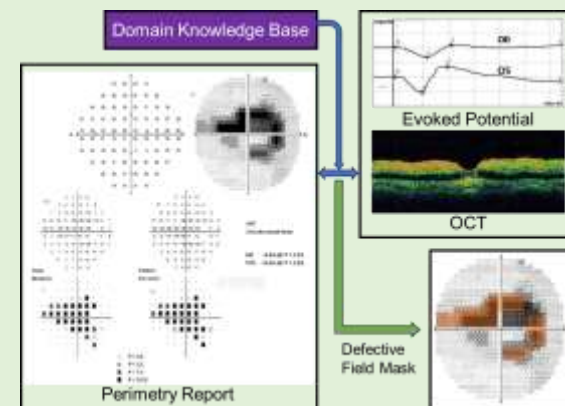
使用者

厨房场景下的物品信息及位置提示

1. 如何识别厨房场景中的物品种类并准确获取物品的方向和距离信息？



基于深度学习和多模态数据的视野缺陷区域分割



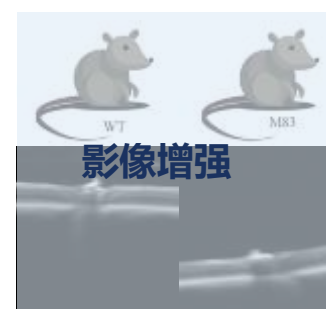
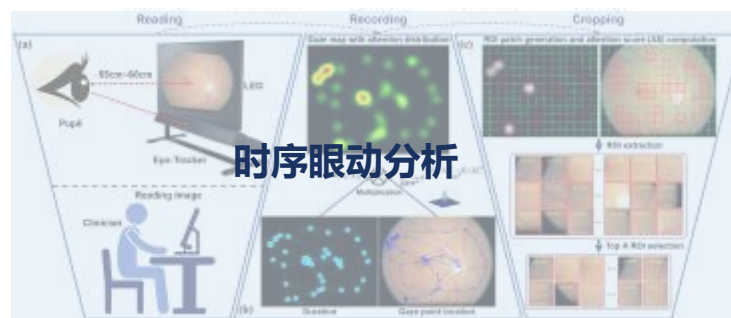
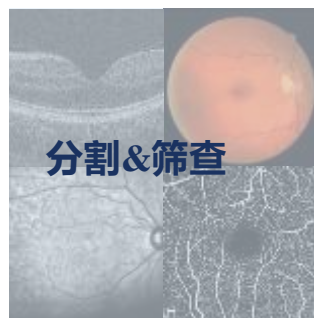
视针对不同类型视野缺损的低视力病人，在视康复中标定视野缺损区域，个性化设计助视方案。

1. 基于目标检测、SAM技术进行眼底神经分割
2. 建立视野计报告和视神经分布的对齐算法
3. 基于孪生神经网络实现多模态视野缺陷区域分割

眼脑联动组

带队人：王星月，刘江

项目内容



可选课题

1. 眼科多模态数据的神经退行性疾病分割和筛查算法研究
2. 帕金森小鼠的眼底OCT影像增强算法研究
3. 时序眼动医学知识驱动的计算机辅助诊断算法研究

超分辨率

transformer

对比学习

注意力

自监督

可信

预期成果



大创项目



学术论文

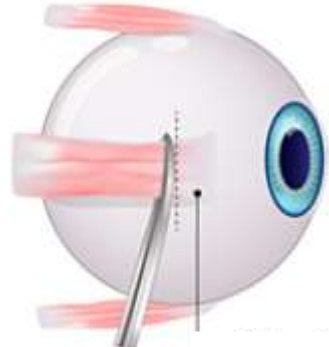
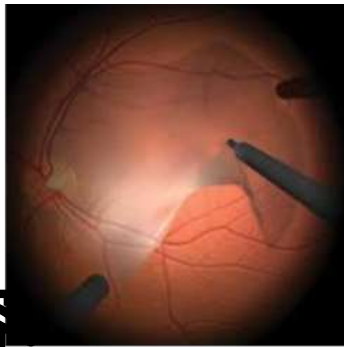
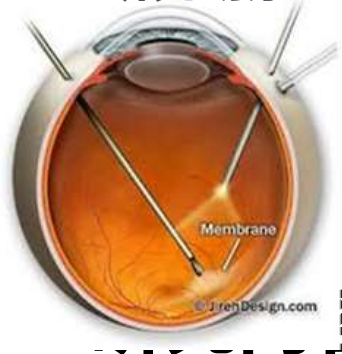


CAD 系统

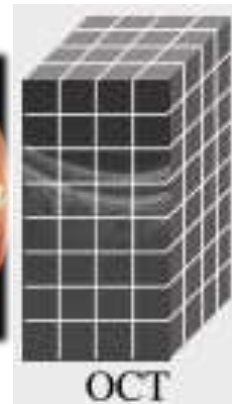
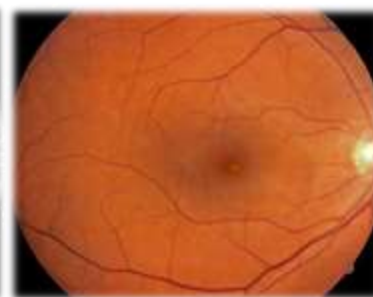
基于图像、视频的医学手术导航研究

带队人：胡衍，刘江

研究场景:



眼科术前多模态影像



眼科术中视频



PP_20850



PP_20875



PP_20900



PP_20925



PP_20950



PP_21100



PP_21125



PP_21150



PP_21175



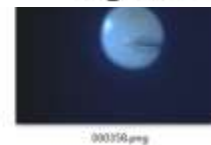
PP_21200



000156.png



000157.png



000158.png



000159.png



000154.png



000160.png

影像三维重建

- 图像分割：小样本、半监督、大模型的眼底影像目标分割
- 多模态图像配准：眼底图像与其他造影图像等的配准
- 二维生成三维：通过眼底图像直接生成三维视网膜结构

视频处理

- 多目标分割跟踪：手术器械及组织的分割与跟踪
- 帧间目标对齐：手术视频或者诊断视频帧间眼睛对齐
- 深度估计：手术器械或者组织的深度信息估计



联系方式

地址：广东省深圳市南山区
学苑大道1088号工学院南楼
445B

联系人：胡老师

邮箱：huy3@sustech.edu.cn

电话：0755-88015223



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--|--|-----------------------|
| <p>Performance Comparison Between Generational and Steady-state Evolutionary Algorithms for Combinatorial Optimisation</p> | <p>Two different Evolutionary Algorithm (EA) models are commonly used: the generational model where all individuals are replaced in each generation (i.e., a non-elitist model) and the steady-state model where only a subset of the population is replaced at every generation (i.e., an elitist model). Recently it has been shown theoretically and experimentally that the steady-state model allows to control the balance between exploitation and exploration of the evolutionary process better than the generational model for Pseudo-Boolean Optimisation. This is achieved by using counter-intuitive selection operators, that yet lead to surprising performance. However, no direct performance comparisons have ever been made between the steady-state EAs using the novel selection operators and traditional Generational EAs. In this project the steady-state and generational models will be compared to evaluate and quantify how well they allow to control the selective pressure of the population. The performance of the algorithms will be evaluated by testing them on hard combinatorial optimisation problems.</p> <p>A high quality project may lead to a published research paper.</p> <p>References <i>Corus, Lissovoi, Oliveto Witt, 2021. On Steady-State Evolutionary Algorithms and Selective Pressure: Why Inverse Rank-Based Allocation of Reproductive Trials is Best</i> https://dl.acm.org/doi/pdf/10.1145/3427474 <i>K. Deb, D. E. Goldberg. A comparative analysis of selection schemes used in genetic algorithms. Foundations of Genetic Algorithms 1, 69-93, 1991.</i> sciedirectassets.com</p> | <p>Pietro Oliveto</p> |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|--|-----------------------|
| <p>Theoretical and Experimental performance analysis of crossover based evolutionary algorithms using inverse selection</p> | <p>Genetic Algorithms (GAs) evolve a population of solutions for optimisation problems by creating new solutions via crossover and mutation in each generation. It is well understood that for the genetic algorithm to make good progress via the crossover operator the population of candidate solutions should be diverse i.e., the solutions should be different.</p> <p>Recently it has been shown theoretically and experimentally how evolutionary algorithms can control the the diversity of the population much better than previously thought. This is achieved by using counter-intuitive <i>inverse selection</i> operators, that yet lead to surprising performance [1]. However, the considered algorithms only use mutation operators (i.e., no crossover), so it is not known to what extent complete genetic algorithms with crossover can take advantage of the greater diversity available.</p> <p>In this project a crossover-based GAs using inverse selection will be analysed both experimentally and theoretically to evaluate how well they can escape local optima when compared against the traditional GAs.</p> <p>A high quality project may lead to a published research paper.</p> <p>References</p> <p>[1] Corus, Lissovoi, Oliveto Witt, 2021. <i>On Steady-State Evolutionary Algorithms and Selective Pressure: Why Inverse Rank-Based Allocation of Reproductive Trials is Best</i> https://dl.acm.org/doi/pdf/10.1145/3427474</p> <p>[2] D.-Dang et.al. <i>Escaping Local Optima With Emergent Diversity</i>. IEEE Xplore Full-Text PDF</p> | <p>Pietro Oliveto</p> |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|---|-------------------|
| Evaluating the performance of artificial immune systems for NP-Hard Number Partitioning Problem | <p>Artificial immune systems (AIS) are a class of bio-inspired algorithms that imitate the natural immune system to solve complex computational problems.</p> <p>In the context of optimisation, AIS mainly differ from more traditional evolutionary algorithms (EAs) by using more sophisticated operators such as hypermutations and ageing.</p> <p>AIS have been proven theoretically to be able to guarantee arbitrarily good approximate solutions for the NP-Hard Number Partitioning problem, while some EAs have been proven to get stuck on bad approximate solutions in the worst case. However, it is unclear which class of algorithms performs better in practice beyond the worst case.</p> <p>The aim of this project is to shed light on this question by comparing experimentally the performance of various AIS and EAs for the NP-Hard Number Partitioning problem and to design improved versions of the algorithms.</p> <p>A high quality project may lead to a published research paper.</p> <p>References:</p> <p><i>D Corus, PS Oliveto, D Yazdani. Artificial immune systems can find arbitrarily good approximations for the NP-hard number partitioning problem. Artificial Intelligence 274, 180-196, 2019</i></p> <p><i>C Witt. Worst-case and average-case approximations by simple randomized search heuristics. Annual Symposium on Theoretical Aspects of Computer Science, 44-56, 2005.</i></p> | Pietro Oliveto |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|---|-------------------|
| Evolutionary Algorithm Design for Permutation-Based Combinatorial Optimisation Problems | <p>Two different Evolutionary Algorithm (EA) models are commonly used: the generational model where all individuals are replaced in each generation (i.e., a non-elitist model) and the steady-state model where only a subset of the population is replaced at every generation (i.e., an elitist model). Recently it has been shown theoretically and experimentally that the steady-state model allows to control the balance between exploitation and exploration of the evolutionary process better than the generational model for Pseudo-Boolean Optimisation. This is achieved by using counter-intuitive selection operators, that yet lead to surprising performance. However, it is unclear whether the results carry over to permutation-based combinatorial optimisation problems such as the travelling salesman problem. In this project steady-state EAs with good exploration/exploitation capabilities for permutation problems will be designed and their performance compared against existing algorithms. A high quality project may lead to a published research paper.</p> <p>A high quality project may lead to a published research paper.</p> <p>References</p> <p><i>Corus, Lissovoi, Oliveto Witt, 2021. On Steady-State Evolutionary Algorithms and Selective Pressure: Why Inverse Rank-Based Allocation of Reproductive Trials is Best</i> https://dl.acm.org/doi/pdf/10.1145/3427474</p> <p><i>Jens Gottlieb, Torben Kruse (2000). Selection in Evolutionary Algorithms for the The traveling salesman problem. Symposium of Applied Computation (SAC 2000).</i></p> | Pietro Oliveto |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--|--|-------------------|
| Time Complexity Analysis of Hyper-heuristics | <p>Rather than manually deciding on a suitable algorithm configuration for a given optimisation problem, hyper-heuristics are high-level search algorithms which evolve the heuristic to be applied. While there are numerous reported successful applications of hyper-heuristics to combinatorial optimisation problems, it is not yet fully understood how well they perform and on which problem classes they are effective. Selection hyper-heuristics (SHHs) employ smart methodologies to select from a pre-defined set of low-level heuristics which to apply in the next decision step. In this project the current foundational understanding of the behaviour and performance of SHHs will be extended by analysing the time complexity of SHHs on unimodal and multimodal optimisation problem classes with significant structures. The analysis will shed light into how and when they can be successfully applied and may lead to insights that allow the development of better performing hyper-heuristics.</p> <p>A high quality project may lead to a published research paper.</p> <p>References <i>Lissovoi, Oliveto, Warwicker, 2019. Simple Hyper-heuristics can control the neighbourhood size optimally for Leading Ones</i> 1801.07546 (arxiv.org) <i>Lissovoi, Oliveto, Warwicker, 2023. When Move-acceptance hyper-heuristics outperform Metropolis and elitist evolutionary algorithms and when not.</i> sciedirectassets.com</p> | Pietro Oliveto |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|--|-------------------|
| Time Complexity Analysis of the limits of steady-state Evolutionary Algorithms and how they can be overcome | <p>Two different Evolutionary Algorithm (EA) models are commonly used: the generational model where all individuals are replaced in each generation (i.e., a non-elitist model) and the steady-state model where only a subset of the population is replaced at every generation (i.e., an elitist model). Recently it has been shown theoretically and experimentally that the steady-state model allows to control the balance between exploitation and exploration of the evolutionary process better than the generational model for Pseudo-Boolean Optimisation. This is achieved by using counter-intuitive selection operators, that yet lead to surprising performance. In this project problem characteristics that lead to poor performance of steady-state EAs will be identified through their time complexity analysis for multimodal optimisation problem classes with significant structures. Then measures will be sought to overcome the identified difficulties. The analysis will shed light into how and when they can be successfully applied and may lead to insights that allow the development of better performing evolutionary algorithms.</p> <p>A high quality project may lead to a published research paper.</p> <p>Reference <i>Corus, Lissovoi, Oliveto Witt, 2021. On Steady-State Evolutionary Algorithms and Selective Pressure: Why Inverse Rank-Based Allocation of Reproductive Trials is Best</i> https://dl.acm.org/doi/pdf/10.1145/3427474</p> | Pietro Oliveto |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|--|-------------------|
| Theory driven design of artificial immune systems | <p>Artificial immune systems (AIS) are a class of bio-inspired algorithms that imitate the natural immune system to solve complex computational problems.</p> <p>In the context of optimisation, AIS mainly differ from more traditional evolutionary algorithms (EAs) by using more sophisticated operators such as hypermutations and ageing.</p> <p>Despite numerous successful applications the theoretical foundations of AIS are still not deeply developed. In this project time complexity analyses of AIS will be performed to shed light on the power of their typical hypermutation and ageing operators in comparison with those employed by more traditional evolutionary algorithms. The gained foundational understanding will lead to the design of improved AIS.</p> <p>A high quality project may lead to a published research paper.</p> <p>References:</p> <p><i>D Corus, PS Oliveto, D Yazdani. Artificial immune systems can find arbitrarily good approximations for the NP-hard number partitioning problem. Artificial Intelligence 274, 180-196, 2019</i></p> <p><i>D Corus, PS Oliveto, D Yazdani. Fast immune-inspired hypermutation operators for combinatorial optimisation. IEEE Transactions on Evolutionary Computation 25 (5) 609-629, 2021</i></p> <p><i>D Corus, PS Oliveto, D Yazdani. When hypermutations and ageing enable artificial immune systems to outperform evolutionary algorithms. Theoretical Computer Science, 832, 166-185</i></p> | Pietro Oliveto |



| Title | Introduction | Supervisor |
|---|---|------------|
| 深度学习基础理论相关： 面向可信模型构建的分布外样本（异常样本输入）检测 | <p>1、研究内容：现多数学习模型都基于封闭世界假设：测试数据抽样自与训练数据相同的分布。当监督学习模型部署于开放世界，测试样本可能来自训练集分布外（out-of-distribution, OOD）。OOD样本检测任务致力于构建可信识别系统，除了能对训练集同分布样本进行预测，也能检测分布外测试样本，予以拒判和交给专业人员进行分析或增量标注。本创新实践课题将在回顾近年来OOD检测工作的基础上，提出新的基于生成式或者判别式模型的异常样本检测方法。尤其是关注于大模型场景下的OOD检测，如保证垂直场景下微调大模型的精度的同时，保持良好的OOD检测性能。</p> <p>2、参考文献：W. Huang et al. Density-driven Regularization for Out-of-distribution Detection. NeurIPS 2022.</p> <p>3、联系人：zhangjg@sustech.edu.cn</p> | 张建国 |
| 深度学习基础理论相关： 面向可解释模型构建的置信度校准研究 | <p>1、研究内容：置信度校准方法旨在校准预测概率，使预测概率能反映真实观测频率。这种方法是提高黑盒模型可解释性的一种重要技术，也是不确定性研究方向中的重要组成部分。近年来，一些研究发现，不同于十年前的神经网络，现代深度学习网络的校准表现很差。通常来说，深度网路输出的概率不能反映真实的概率，倾向于提供过置信的判别概率。而最近也有很多经验性的矫正方法被提出，然而仍然还有很多本质的问题没有被解决，例如，理论上来说什么样的校准函数是较优秀、最优的矫正函数，如何从理论出发设计校准优化目标等。本创新实践课题，将在学习、回顾置信度校准发展的基础上，提出具有良好解释性且表现优异的矫正方法，以及探究大模型相关的置信度校准评价与学习策略。</p> <p>2、参考文献：Guo C, Pleiss G, Sun Y, et al. On calibration of modern neural networks, ICML, 2017.</p> <p>3、联系人：zhangjg@sustech.edu.cn</p> | |



| Title | Introduction | Supervisor |
|---------------------------------------|--|------------|
| 深度学习理论相关： 重播式连续学习的样例选择方法研究 | <p>1、深度学习模型在训练数据分布改变时，会出现灾难性遗忘现象，也就是学习解决新任务的能力会覆盖掉解决旧任务的能力。连续学习（Continual Learning）旨在缓解这种遗忘现象。重播式方法（Replay-Based Method）是连续学习的一类有重要影响力的方法，通过将每个任务的少量数据作为样例保存，在训练新任务时融入这些样例。本课题拟调研重播式方法中的多种选取样例的方法，分析其在不同连续学习场景中的效果，并依此提出新的样例选择方法。</p> <p>2、参考文献：Van de Ven G M, Tolias A S. Three scenarios for continual learning[J]. arXiv preprint arXiv:1904.07734, 2019.</p> <p>3、联系人：zhangjg@sustech.edu.cn</p> | 张建国 |
| 深度学习语义分割相关： 面向开放词汇语义分割的多模态模型适配方法研究 | <p>1、研究内容：语义分割是计算机视觉领域中重要的视觉感知与理解任务之一。传统的语义分割任务通常限制在封闭的词汇范围内，难以适应开放场景中的丰富语义类别。随着多模态视觉语言预训练模型（VLMs）的发展，开放词汇语义分割开始通过利用VLMs的图像与文本对齐能力，根据给定的任意文本描述来匹配图像中的语义目标，从而在一定程度上实现对广泛语义类别的识别。然而，现有的VLMs大多依赖于大规模图像级数据进行预训练，其在语义分割这类区域级密集预测任务中的表现仍有局限。本研究课题旨在探索现有VLMs的高效适配方法，通过在有限的数据集上训练，以达到更优秀的零样本语义分割性能。</p> <p>2、参考文献：Wu J, Li X, Xu S, et al. Towards open vocabulary learning: A survey. TPAMI 2024.</p> <p>3、联系人：zhangjg@sustech.edu.cn</p> | |



| Title | Introduction | Supervisor |
|-----------------------------|--|------------|
| 深度学习目标检测相关：面向复杂场景的小目标检测方法研究 | <p>1、 研究内容：在深度学习领域中，小目标检测是目标检测任务中的常见场景和重要难点问题，本课题旨在通过对先进的目标检测算法进行研究，以增强模型对于小目标的检测和识别能力。</p> <p>2、 参考文献：Liu Y et al. A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection. Expert Systems with Applications 2021.</p> <p>3、 联系人： zhangig@sustech.edu.cn</p> | 张建国 |
| 深度学习目标检测相关：开放世界的零样本目标检测方法研究 | <p>1、 研究内容：当前深度学习算法大多依赖于在任务前期对待检测目标进行类别定义，并进行人工数据标注。此类方式检测的目标类别有限，无法检测未定义的类别目标。本课题旨在通过对跨模态方法和目标检测方法的研究，实现对开放世界的零样本目标检测方法研究。</p> <p>2、 参考文献：A. Radford et al. Learning Transferable Visual Models from Natural Language Supervision. ICML 2021.</p> <p>3、 联系人： zhangig@sustech.edu.cn</p> | |
| 深度学习目标检测相关：基于目标检测的域适应方法研究 | <p>1、 研究内容：域适应是计算机视觉领域中较为活跃的研究方向，旨在将模型从一个领域（称为源域）迁移到另一个领域（称为目标域），从而提高模型在目标域上的性能。本课题拟对目标检测的域适应方法进行研究，期望能够有效增强检测模型的泛化性能，降低模型在新域上的训练成本，同时减少模型对数据标注的需求。</p> <p>2、 参考文献：P. Oza et al. Unsupervised Domain Adaptation of Object Detectors: A Survey. TPAMI 2023.</p> <p>3、 联系人： zhangig@sustech.edu.cn</p> | |



| Title | Introduction | Supervisor |
|--------------------------------------|--|------------|
| 大模型建模理论相关： (基座大模型评估方法研究) | <p>1、 研究内容：在2020年，千亿级别的通用大模型GPT-3发布后，其卓越性能引起了广泛关注。类似的通用基座大模型如雨后春笋般涌现。随后，OpenAI率先发布了大模型下游应用产品：ChatGPT。这一示范性应用的推出，激发了许多专业领域研发专用大模型的热情。这些专用模型被称为垂直大模型。从众多大模型中挑选适合的模型作为基座，是构建垂直大模型的关键步骤。然而，目前的选择方法存在较强的主观性和局限性。针对这一问题，本课题将在回顾过往大模型评估方法的基础上，提出新的客观评估方法。</p> <p>2、 参考文献：Wenjing Yue et al. TCMBench: A Comprehensive Benchmark for Evaluating Large Language Models in Traditional Chinese Medicine. ArXiv:2406.01126</p> <p>3、 联系人：zhangjg@sustech.edu.cn</p> | 张建国 |
| 大模型建模理论相关： (医疗垂直大模型建模研究) | <p>1、 研究内容：在2022年，OpenAI利用其在大模型研究领域的领先优势，推出了首个大模型下游应用产品：ChatGPT。该示范性产品的成功推出，极大激发了医学领域对于开发专用大模型的热情，这类模型被称为垂直大模型。在垂直大模型的开发过程中，涉及到增量学习、人类偏好对齐以及分布式推理等多种高级微调技术。目前，医疗垂直大模型的研发主要由人工智能专家团队主导，但这种做法存在局限性，导致模型难以精确满足实际的临床需求，限制了其落地应用。针对这一问题，本课题组已组织了一支由资深医疗专家构成的团队，计划深入研究并开发更加贴合临床需求的医疗垂直大模型。</p> <p>2、 相似的模型实例：https://www.huatuogpt.cn/#/</p> <p>3、 联系人：zhangjg@sustech.edu.cn</p> | |



| Title | Introduction | Supervisor |
|------------------------------|---|------------|
| 大模型建模理论相关： (医疗大模型开源挑战赛) | <p>1、在医疗大模型开源挑战赛汇总网站中，研究者精选了一系列具有重大科学探究意义的挑战赛。这些赛事通过统一的开源数据和评估标准，比较各种算法的性能，每年吸引全球众多国家和地区的研究人员参与。开源赛事的举办不仅推动了新算法的发展，也促进了医学与人工智能技术的深度融合。选择加入我们的创新实践课题，你将从最新发布的挑战赛中挑选感兴趣的项目。在与指导老师的紧密沟通和确认后，你将参与这些激动人心的比赛，并在赛后撰写研究论文，展示你的科研成果和技术创新。</p> <p>2、挑战赛汇总网站：https://github.com/FreedomIntelligence/Medical_NLP</p> <p>3、联系人：zhangjg@sustech.edu.cn</p> | 张建国 |
| 计算机视觉和机器学习中的学生自选课题 | <p>1、同学根据兴趣自行拟定计算机视觉、机器学习等领域的实践项目，具体创新实践项目题目的拟定、开展与实施需要先和指导老师确认。</p> <p>2、联系人：zhangjg@sustech.edu.cn</p> | |



| Title | Introduction | Supervisor |
|-----------------------|--|------------|
| 基于自监督方法的自动驾驶场景间域适应性研究 | 数据集由于采集平台、设备或场景不同、图片分辨率不同、城市的车辆密度不同等，深度学习模型存在域适应性问题。明显地，基于数据驱动的自动驾驶方法普遍存在在某个数据集上训练的模型，换个应用场景，模型精度极具下降。本项目基于BEV目标检测任务，研究不同自动驾驶场景数据间域适应性的自监督/弱监督学习方法，提高感知模型在目标域的鲁棒性。 本目前期研究工作已发表一篇CCF-B会议。目标：发表科研论文一篇。 | 郝祁，张美莹 |
| 自动驾驶关键场景数据挖掘及相似性评估 | 利用真实路采驾驶数据集（相机，激光雷达、IMU等），针对不同粒度（比如，粗粒度-自车变道，细粒度-自车变道前方有障碍）场景，通过基于规则和基于模型的方法从长时间数据中提取关键场景。除此之外，针对道路、环境、交通参与者、主车交互等建立场景的图模型，通过度量特征间的距离对这些关键场景进行相似度分析。 本项目正在与企业进行项目合作，可提供到企业实习机会。目标：发表专利/科研论文一篇。 | |
| 构建高相似度动态场景下驾驶行为通用评价方法 | 针对细粒度场景（比如，自车变道前方有障碍，十字路口直行通过遇横向车辆等）中相似/同类场景，从理论上建立一套评价特定驾驶行为优劣的机制，包括评价驾驶行为的安全、舒适、快捷、合理性等方面的表现，并构建长时间驾驶（包含多个细粒度场景）下驾驶行为评价方法。 本项目正在与企业进行项目合作，可提供到企业实习机会。目标：发表专利/科研论文一篇。 | |
| 数据驱动的自动驾驶场景重建及动态场景生成 | 常见的自动驾驶仿真器（如 Carla）基于手工搭建的场景，在场景的真实性方面难以得到保证，并且与真实数据间存在很大的差异性。本项目基于 数据驱动的方法对真实的场景进行重建，结合Unreal5 游戏引擎，在重建的静态场景基础上加入动态物体，生成自动驾驶数据流，获取更加真实的仿真数据。 本目前期静态场景重建已投稿，计划进一步深入研究动态场景生成。目标：发表专利/科研论文一篇。 | |



| Title | Introduction | Supervisor |
|------------------------|---|------------|
| 基于大模型的自动驾驶场景生成技术研究 | 随着 AIGC (AI-Generated Content)技术的发展，大型模型在自动驾驶场景生成中的应用主要体现在模拟训练方面。通过构建出多样化的道路和气候条件，大型模型能够模拟出各种复杂的交通场景，帮助自动驾驶系统在面对真实环境时能够作出更加准确和安全的决策。本项目基于Diffusion transformer架构，研究从文本语义到自动驾驶图像/视频场景生成方法，实现构建合理的且可用于数据增强的驾驶场景。 本项目有初步的研究基础，计划进一步探索。目标：发表专利/科研论文一篇。 | 郝祁，张美莹 |
| 基于大模型的自动驾驶关键场景检索增强技术研究 | 检索增强生成RAG (Retrieval Augmented Generation) 是2023年基于LLM系统中最受欢迎的架构，通过自有垂域数据库检索相关信息，然后合并成为提示模板，给大模型生成漂亮的回答。本项目基于RAG架构中检索阶段，将输入的不同模态（文本或图像）数据，利用LLM生成多个查询（query），通过这些查询让相关各个方面在上下文显现出来，之后使用生成的查询进行向量搜索，基于其结果集中的显示方式对自动驾驶场景进行内容定位。 本项目与华为进行项目合作。目标：发表专利一篇。 | |
| 大模型强化学习策略优化技术 | 基于人类反馈的强化学习（RLHF）的大模型微调，能保证大模型输出与人类的期望和偏好一致，避免产生有害或不良输出。但是，大模型强化学习由于其本身训练过程的复杂性（涉及多模型交互、词表方差大等），造成训练过程很不稳定。本项目基于强化学习策略优化方法，将action转换为词表单词输入，实现大模型的微调训练，并研究有效提升稳定性方法。 本项目与华为进行项目合作。目标：发表专利一篇。 | |



| Title | Introduction | Supervisor |
|---------------------------|--|------------|
| 特斯拉纯视觉方案自动驾驶的深度占用网络感知技术研究 | <p>背景：特斯拉最新提出的占用网络（Occupancy network）感知技术展示了领先的感知能力，是目前自动驾驶感知领域的最新研究热点。其深度占用网络关注通用几何信息，不过度区分语义，更体现空间是否“被占用”，很好地解决自动驾驶场景中截断目标、形状不规则、未有清晰语义（比如挂车、树木、垃圾、以及石子等）的目标检测问题。</p> <p>目标：本课题研究SOTA的占用网络模型在自动驾驶不利场景中的问题及优化，在实验室GPU计算集群上训练深度占用网络模型，在仿真和真实自动驾驶中部署和测试占用网络感知，实现高效的占用网络感知深度模型。相关研究成果与导师团队合作发表文章。</p> <p>已有研究：研究小组已经采集用于训练占用网络的数据，并训练和测试了多个占用网络模型，为本课题的开展提供了很好的先验性研究。</p> | 郝祁，兰功金 |
| 4D毫米波雷达在自动驾驶深度占用网络感知中的作用 | <p>背景：4D毫米波雷达及相关深度网络算法在自动驾驶领域正在革命性兴起。解决了3D毫米波雷达识别静止目标困难、横向移动检测不足、缺乏高度信息、穿透力不足、分辨率较低、延迟大等多种局限，且相比激光雷达具有价格低，稳定性高等特点，其与视觉占用网络感知融合的深度网络模型方案正成为当前自动驾驶的主流方案。</p> <p>目标：本课题研究4D毫米波雷达与视觉融合的深度占用网络感知技术，在GPU计算集群上训练多模态的深度占用网络模型，并在仿真和真实自动驾驶中部署和测试占用网络感知，实现4D毫米波雷达与视觉融合的占用网络感知。相关研究成果与导师团队合作发表文章。</p> <p>已有研究：研究小组已经初步调研了相关研究，并完成了多个视觉占用网络模型的训练和测试。本课题具有很好的先进性和可行性。</p> | |



| Title | Introduction | Supervisor |
|---------------------|--|------------|
| 端到端深度网络模型的自动驾驶技术研究 | <p>背景：2024年特斯拉推出的FSD V12端到端自动驾驶方案展示了令人震撼的智能程度，掀起了端到端自动驾驶技术的热潮，国内国际各大自动驾驶公司也纷纷跟进布局端到端的自动驾驶技术研究。</p> <p>目标：本课题在实验室已有端到端自动驾驶研究的基础上，进一步探索端到端自动驾驶技术的性能，特别是在处理极端场景时的性能表现，在实验室GPU计算集群上实现端到端模型的训练及性能测试，实现端到端模型的优化和部署。相关研究成果与导师团队合作发表文章。</p> <p>已有研究：研究小组已经调研了端到端自动驾驶相关深度网络模型，并初步开展了相关方案的可行性分析，为本课题的可行性提供了很好的支撑。</p> | 郝祁，兰功金 |
| 自动驾驶多模态深度占用网络感知技术研究 | <p>背景：特斯拉最新推出的占用网络（Occupancy network）感知技术展示了领先的感知能力，但纯视觉的占用网络感知对距离的估计普遍存在稳定性和精度问题，此类问题在地下停车场等场合显得尤为突出。多模态（视觉，激光雷达，毫米波雷达）融合的占用网络感知技术可以很好的解决此类问题。</p> <p>目标：本课题拟训练和测试当前SOTA的多模态深度占用网络感知模型，优化模型在自动驾驶不利场景中的稳定性和精度，并在地下停车场仿真和真实环境中测试。相关研究成果与导师团队合作发表文章。</p> <p>已有研究：研究小组已经调研了多模态深度占用网络感知相关研究工作，并对不同数据集有深入分析，团队训练了多个占用网络深度模型，已有一辆多模态自动驾驶测试车辆（比亚迪）以及工学院地下停车场1：1仿真3D模型，本课题具有很好的可行性和先进性。</p> | |



| Title | Introduction | Supervisor |
|----------------|--|------------|
| 大模型的自动驾驶不利场景生成 | <p>背景：以特斯拉端到端（内含占用网络）为代表的自动驾驶技术，其挑战主要是对高质量自动驾驶长尾数据和不利测试场景采集的要求，当前的绝大多数训练数据和测试场景已经很难用于提升自动驾驶性能。自动驾驶不利场景的生成，可以提供高质量的长尾数据以及测试场景。以ChatGPT为代表的大模型可以为自动驾驶场景生成提供场景语义理解和多样性，正在成为自动驾驶领域的研究热点。</p> <p>目标：本课题拟调用通用大模型接口，探索大模型的自动驾驶不利场景生成，提供高质量的自动驾驶长尾数据及测试场景。用于训练和测试端到端的自动驾驶模型。相关研究成果与导师团队合作发表文章。</p> <p>已有研究：研究小组已经实现了基于强化学习的自动驾驶不利场景生成，并调研了当前大模型的自动驾驶场景生成技术，初步形成课题方案，本课题具有很好的可行性，以及突出的先进性。</p> | 郝祁，兰功金 |



| Title | Introduction | Supervisor |
|--------------------|---|------------|
| 交通大数据下的出行数据分析研究 | 交通大数据下的出行数据分析研究是一个涉及多学科交叉的领域，它利用大数据技术对交通出行数据进行收集、处理、分析和应用，以揭示交通出行的规律和特点，为交通规划、管理和服务提供科学依据。 | 张凌宇/宋轩 |
| 基于大语言模型的网约车需求量预测研究 | 基于大语言模型的网约车需求量预测研究，主要利用大语言模型对海量文本数据的理解和分析能力，结合深度学习技术，如卷积神经网络（CNN）、长短期记忆网络（LSTM）和图神经网络（GNN）等，对网约车需求量进行建模和预测。 | |
| 智慧交通中的乘客出行模式研究 | 智慧交通中的乘客出行模式研究是一个复杂而多维的领域，它涉及对乘客出行行为的深入理解、数据分析以及技术应用等多个方面。随着城市化进程的加快和科技的飞速发展，乘客的出行模式正在经历深刻的变化。传统上，公共交通、私家车、步行和骑行是主要的出行方式。然而，在智慧交通的推动下，新的出行模式不断涌现，如共享单车、网约车、无人驾驶出租车等。这些新模式不仅丰富了乘客的出行选择，还提高了出行的便捷性和灵活性。 | |



| Title | Introduction | Supervisor |
|--|---|------------|
| 基于强化学习的海上船舶管理航线规划 | 本研究旨在开发一种创新算法，利用强化学习技术优化海上船舶的航线规划。通过解决海上运输中的航线规划挑战，所提出的方法提高了导航效率，减少了燃料消耗，缩短了旅行时间，并降低了碰撞风险。通过模拟各海域的环境变化，评估碰撞概率，并考虑船舶的操作状态，该方法动态调整航线，确保安全高效的导航。该方法的自适应特性确保船舶能够避开潜在的危险和其他船只，在复杂多变的海上环境中更加安全高效地航行。本研究不仅对海上运输业具有重要的实际价值，还为智能交通系统的发展提供了新的见解和方法。 | 范子沛/宋轩 |
| 探究用户信息对大模型预测POI的影响以及如何使大模型在不同用户群体间实现预测的公平性 | 在进行POI prediction时（下一个兴趣地点预测时），不同偏好的用户，他们的兴趣地点也会不同。在实际情况中，用户的兴趣偏好也会影响模型对兴趣地点的预测，为了探究用户的偏好会对大模型的预测产生多大的影响，我们在训练过程中通过用户的历史轨迹数据推导出用户的人设，如下图，然后使用推导出的人设在测试集预测时加入人设的信息，1.查看大模型的预测结果相比之前有什么区别；2.并探究不同人设之间的预测准确度差异，如何修改prompt可以使得模型对于不同人设的预测准确度尽可能相等（使模型的预测结果在不同人设之间更公平） | |
| 基于负迁移抑制的POI预测的群体公平性 | 将POI数据集中的用户进行分组（可以根据用户的性别、年龄、社交网络上的朋友数量或者历史轨迹数量进行多属性的分组）。由于每个组的数据在数据集上存在数据不平衡的问题，所以最终分组后的POI预测准确度存在不公平的问题，为了解决这个问题，可以将这些分组的用户的POI prediction看作是一个多任务的问题，使用负迁移相关的技术尽可能的改善最终分组后POI预测的不公平性。 | |



| Title | Introduction | Supervisor |
|---------------|---|------------|
| 基于深度学习的动态图建模 | <p>图，作为描绘真实世界错综复杂关系的工具，被广泛应用于社交网络、自然语言处理、计算机视觉、计算金融和推荐系统等领域。图机器学习在节点分类、链路预测和图聚类等任务上取得了巨大成功。然而，大多数图机器学习方法假设图是静态的，这限制了其在动态变化的现实世界中的应用。动态图学习（Temporal Graph Learning），通过将时间信息与图的特征和结构联合建模，显著提升了其在真实世界中的预测能力。随着人工智能、大数据等技术与动态图学习的深度融合，模型在推荐系统、舆情监控、传染病预防、欺诈检测等应用上取得了显著进展。本项目旨在探究不同深度学习框架在动态网络建模中的优劣，并结合推荐系统或社交网络等实际任务，分析动态网络演化的本质。</p> | |
| 基于深度学习的股票价格预测 | <p>股票市场风险与收益并存，是全球金融系统的重要组成部分；但由于经济指标、财务报告、全球新闻等各个方面动态且复杂的影响，预测股票价格一直是一个极具挑战性的问题。近年来，深度学习凭借其捕获股价时间维度特征的强大功能得到研究人员们的青睐，然而日益复杂的网络架构在性能提升方面收效甚微，且往往受到有限的股票数据的影响。因此，我们跳出股票价格预测在时间维度的局限，关联多只股票，通过对复杂的相关性建模来提高预测的准确度。Transformer模型的自注意力机制可以有效结合多只股票，学习股票间不对称的、动态变化的关系；也可以在时间维度学习股票价格长期的变化模式，有效对复杂的股票市场建模。我们的实验基于Microsoft Qlib平台，集成了庞大的中国/美国股价历史数据和广泛的股价预测基准模型。通过整合Transformer模型在建模时间/多只股票相关性的功能，我们能够有效提升股价预测准确率，最大化股票投资收益。</p> | 姜仁河/宋轩 |



| Title | Introduction | Supervisor |
|----------------|---|------------|
| 基于深度学习的风力发电预测 | <p>风力发电作为可再生能源的重要组成部分，在全球能源系统中占据着越来越重要的地位。然而，由于风力发电受气象条件、地理环境等多种复杂因素的影响，预测风力发电量一直是一个极具挑战性的问题。近年来，深度学习凭借其强大的建模能力，得到了研究人员的广泛关注。随着图学习和时空预测领域的快速发展，最近的研究致力于突破风力发电预测在时间维度上的局限，通过建模不同风力发电站之间的相关性、引入气象数据、分析气候变化趋势等多种方式来辅助风力发电预测。我们的实验基于真实的风力发电数据，集成了环境，风速以及引擎相关的辅助数据，并采用最新的时空预测模型。通过有效地挖掘气象数据的历史规律，建模风力发电站之间的复杂关联性，从广泛的气象报告和气候变化研究中分析趋势，从而显著提升风力发电预测的准确率，最大化风力发电的利用效率。</p> | 姜仁河/宋轩 |
| 基于深度学习的多变量时序预测 | <p>多变量时序预测是指利用时间序列数据中多个变量的信息，通过建立复杂的关联模型来预测未来的趋势和变化。在实际应用中，例如气象预测、股市走势预测以及工业生产规划中，多变量时序预测能够提供更加全面和准确的预测结果，帮助决策者制定更有效的策略和措施。基于深度学习的方法在多变量时序预测中展现出了显著的优势。相较于传统的统计模型，深度学习模型能够处理复杂的非线性关系和大规模数据，具备更强的表达能力和预测精度。通过结合卷积神经网络（CNN）、循环神经网络（RNN）及其变种，以及注意力机制等最新技术，我们能够有效地捕捉多变量之间的时空依赖关系，从而提升预测模型的准确性和鲁棒性。本课题旨在探索基于深度学习的多变量时序预测方法，结合多个真实的多变量时序数据集，以提高预测精度和实用性，设计出简洁有效的模型。</p> | |



| Title | Introduction | Supervisor |
|------------------------|--|------------|
| 多模态学习驱动的视频生成模型研究及应用 | <p>实验旨在探索跨模态学习方法，实现从静态照片图像到连续动态视频的转换。通过学习了解几种主要深度学习模型如生成对抗网络（GANs）、变换器（Transformer）和变分自编码器（VAEs）、扩散模型（Diffusion Models）等，研究如何有效学习并整合不同模态数据的潜在表示，并通过学习图像序列之间的动态关系，开发能够生成连贯、逼真视频模型。该研究不仅探索技术创新，还探讨其在历史照片视觉重现、动画动漫等领域的潜在应用。已有一位同学。</p> | 宋轩/舒襄 |
| 基于生成式大模型的人工智能体语言行为模型研究 | <p>本实验旨在构建基于生成式大模型的人工智能体语言行为模型，探索如何优化智能体自主决策能力、生成能力，语言表达和行为控制能力，强化NPC的智能化水平。智能体可以在某一类小说背景下（如《西游记》、《三体》、《三少爷的剑》），自定义动态生成忠于原故事世界观和风格的剧情、角色互动和支线任务，重点在于如何控制大模型的幻觉，使智能体生成的支线剧情和主线剧情改动保持在可控范围内；并根据玩家的选择和特定情境下表现出符合角色、更真实和复杂的行为模式；为现代叙事类游戏创造丰富、沉浸和互动的游戏体验。已有一位同学。</p> | |
| 基于位置感知技术与AIGC的随境游戏引擎 | <p>随境游戏（Pervasive Game）通过位置感知技术让线上游戏情节与线下环境内容相融合，为玩家探索游览线下场景、沉浸式游戏与社交互动提供一种新型的体验形式。在本创新实验中，我们将基于Unreal或Unity游戏开发引擎，探索蓝牙低功耗（BLE）技术如何检测玩家移动设备与线下游戏对象（如NPC）的距离或相对位置，从而实时更新游戏状态。另一方面，我们将结合对话式AI、文生图模型生成、调整游戏内容。实验的探索将逐步导向一个基于蓝牙接触追踪技术与AIGC的通用随境游戏引擎，可用于开发任意的随境游戏脚本。</p> | |



| Title | Introduction | Supervisor |
|-----------------------|---|------------|
| 基于大模型的调度优化方法研究 | 本研究致力于探索一种基于大模型的调度优化方法。通过运用大模型微调和对齐技术，对调度问题进行建模和分析，以实现对资源的高效配置和任务的优化调度。旨在提高调度效率、降低成本、提升管理水平。 | 宋轩 |
| 基于大模型的汽车检修生命周期价值链协同模型 | 本研究主要聚焦于构建一个基于大模型的汽车检修生命周期价值链协同模型，旨在通过大数据分析和智能化决策，实现汽车整个检修生命周期中各个环节的高效协同。通过优化资源配置、提高生产效率、降低维护成本，从而提升整个汽车价值链的核心竞争力。 已有二位同学。 | |
| 基于大模型的制造业产品价值链决策分析系统 | 本研究旨在构建一个基于大模型的制造业产品价值链决策分析系统。该系统利用大模型自动解析数据分析需求，并生成相应的SQL查询语句，以实现对制造业产品管理的智能决策支持。研究目的是希望有效提高制造业企业的运营效率、降低成本、优化产品价值链，为企业提供科学、精准、实时的决策依据。 已有二位同学。 | |
| 基于大模型的制造业产品生命周期因果推断方法 | 本研究主要探索一种基于大模型的制造业产品生命周期因果推断方法。通过结合大模型和因果推断技术，研究产品从设计、生产、销售到报废的整个生命周期中各因素之间的因果关系，挖掘关键因素对产品生命周期的影响，为企业优化产品设计、提高生产效率、延长产品使用寿命提供科学依据。 已有一位同学。 | |
| 基于大模型的汽车检修价值链专家知识系统 | 构建一个基于大模型的汽车检修价值链专家知识系统。通过结合大模型和向量数据库，将汽车检修领域的专业知识与计算模型相结合，实现对汽车检修全过程的智能化决策支持。旨在提高汽车检修效率、降低成本、提升服务质量，为汽车检修提供科学、精准、实时的决策依据。 | |
| 智慧海事 | 海上运输作为国际贸易和全球经济的重要组成部分，支撑了国际贸易的大量货物流动，而且对全球供应链的高效运作起着至关重要的作用。然而，随着全球化贸易的增长和国际市场的扩展，海上运输面临着越来越多的挑战，如环境保护、航运安全、以及航道和港口的拥堵问题。 根据国际海事组织（IMO）的规定，载重量在300吨以上的船舶和所有客船，都需要安装船舶自动识别系统（AIS）。AIS和其他航行安全技术的应用，提高了船只航行的安全性和效率。 智慧海事命题旨在通过AIS数据让我们更好的理解海上交通的流动方式，并进一步对航运安全、海上交通管理、环境监测和航运物流等领域展开研究。 | |



| Title | Introduction | Supervisor |
|---------------------------------------|--|---|
| <p>应用开发类课题</p> <p>面向海量数据的散点图可视化平台</p> | <p>本课题将设计并实现一套基于浏览器的散点图可视化平台，支持十万至百万数据点量级散点图的交互式分析。</p> <p>要求：对前端开发有兴趣，最好有前端开发经验。</p> | |
| <p>应用开发类课题</p> <p>NSGA-II算法可视化平台</p> | <p>本课题将设计并实现一套面向NSGA-II演化多目标优化算法的可视化平台，用户可以使用该平台探索算法演化过程和结果的细节内容，并支持对优化过程的交互式诊断和干预。</p> <p>要求：对前端开发有兴趣，最好有前端开发经验。</p> | |
| <p>研究类课题</p> <p>基于增强现实的海报展示</p> | <p>本课题主要探究如何在物理海报上添加AR效果，实现信息增强。主要实现平台为浏览器端和移动设备。</p> <p>要求：对可视化相关的科研有兴趣</p> | |
| <p>设计实现类课题</p> <p>数据可视化作品设计</p> | <p>数据可视化常被用来发现数据中的模式、规律、异常。在本课题中，你可以使用适合的数据可视化方法，展示数据中所包含的信息和潜在模式，讲述数据背后的故事。数据题材不限，可以是个人爱好（诗词歌赋，电影游戏），可以是每日生活（南科今天发生了什么？），也可以放眼世界（时事动态、史海钩沉），只需要你能获得到一定量的有效数据，以及拥有一颗渴望表达的心。最终作品呈现方式可以是静态的信息图，也可以是交互式可视化系统，甚至是可互动的三维世界。</p> <p>一些可以参考的作品集：</p> <ul style="list-style-type: none"> - 澎湃号“一头倭瓜精”（某数据新闻研究者的作品收藏）（https://www.thepaper.cn/sparker_3178232） - 信息图作品《宋词缱绻，何处画人间》（http://fms.news.cn/swf/2018_sjxw/quansongci/index.html#/） - 数据新闻作品《周琦的NBA之路》（http://www.cad.zju.edu.cn/home/vagblog/?p=5395） | <p>马昱欣</p>  |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|---|----------------------------------|
| AR Glasses based real-time classroom data visualization | <p>Developing in Unity a demo application for AR glasses which displays data gathered from a smart classroom.</p> <p>The data are already gathered for testing this demo, which need to be rendered as if they are in real-time during a class session. Wearing the glasses, a user should see the displayed data being spatially distributed and oriented in different locations of the space, using the spatial computing capability of the AR glasses.</p> | 安鹏铖 (设计学院) & 马昱欣 |
| Enhancing Body Ownership Using Haptic Retargeting for Immersive Virtual Reality Experience | <p>We are looking to explore the potential of haptic retargeting in enhancing body ownership in virtual reality.</p> <p>Haptic retargeting allows a single physical object to provide touch feedback for different virtual objects. This is achieved by using visual dominance over proprioception to separate the real and virtual hands. Students will develop the application using a tracking system (e.g., Optitrack), an HMD (e.g., VIVE), and Unity.</p> <p>Additionally, they may need to learn HCI research methods.</p> | Seungwoo Je (设计学院) & 马昱欣 |
| 参与其他科研项目 | 当面详谈 | 马昱欣 |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|-------------------------|--|-------------------|
| 针对机器人控制任务的多目标演化算法设计 | 本项目将开发和设计张量化的多目标演化算法，专门用于处理多目标机器人控制任务。同时，本项目还将探索不同的控制策略结构，研究其如何影响机器人的控制效果。 相关开源项目: https://github.com/EMI-Group/evox | 程然 |
| 针对机器人控制任务的类脑神经网络设计 | 基于谷歌JAX框架，从神经网络结构或训练方法的角度研究与设计用于机器人控制任务的类脑算法。 | |
| GPU加速的超大规模组合优化算法库研发（姜涛） | 基于EvoX框架，本项目将开发和设计针对超大规模组合优化问题的算法库。该项目包含两个核心部分：首先，针对不同类型的经典组合优化问题，进行张量化的问题设计与开发，实现超大规模问题的快速评价；其次，利用GPU加速技术，设计和开发出能够高效求解组合优化问题的算法。 相关开源项目: https://github.com/EMI-Group/evox | |
| GPU加速的高并发演化强化学习算法库研发 | 使用谷歌JAX算法库在GPU上实现高并发模拟环境下的端到端演化强化学习算法库研发，目前主体框架已完成。 | |
| 大模型演化融合算法设计 | 对现有开源大语言模型进行模型融合，得到结合原有模型的优势的语言模型。详情可参照论文 Evolutionary Optimization of Model Merging Recipes (https://arxiv.org/abs/2403.13187)。 | |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--------------------|--|-------------------|
| OpenCV development | OpenCV is an open-source library for computer vision, and has been widely used in research and industry. The students are expected to contribute some algorithms to OpenCV on deep learning, 3D point cloud and others. The students should be good at C++ programming. | 于仕琪 |



| Title | Introduction | Supervisor |
|---|---|------------|
| 基于编码的后量子密码学 Code-based post-quantum cryptography | <p>随着量子计算的快速发展，基于大整数分解和离散数学的公钥密码机制面临巨大挑战。基于编码的后量子密码学是最具潜力的技术路线之一，本项目将研究在NIST第四轮讨论的BIKE等候选后量子密码标准。主要参考文献为：</p> <p>T. Wang, A. Wang, and X. Wang, Exploring decryption failures of BIKE: new class of weak keys and key recovery attacks. CRYPTO 2023.</p> <p>V. Weger, N. Gassner, and J. Rosenthal, A survey on code-based cryptography, arXiv: 2201.07119v4, 2024.</p> | 王琦 |



| Title | Introduction | Supervisor |
|--|--|------------|
| 保护隐私计算协议的应用与实现 Application and Implementation of privacy-preserving computing protocols | <p>本项目尝试将适用于最新场景的安全多方计算协议应用于两方协作数据智能处理，并探索优化落地于相关具体场景。</p> <p>主要参考文献：</p> <p>[1] Y. Liu, Q. Wang, and S.M. Yiu, Making private functions evaluation safer, faster and simpler. In Proceedings of The 25th International Conference on Practice and Theory of Public-Key Cryptography (PKC 2022), Yokohama, Japan, Mar. 7-11, 2022.</p> <p>[2] Y. Liu, Q. Wang, and S.M. Yiu, Blind polynomial evaluation and data trading. In Proceedings of The 19th International Conference on Applied Cryptography and Network Security (ACNS 2021), online, Kamakura, Japan, Jun. 21-24, 2021.</p> | 王琦 |



| Title | Introduction | Supervisor |
|-------------|---|------------|
| 用户认证的安全漏洞研究 | <p>本项目将探索验证现实生活中用户认证存在的安全漏洞，并尝试给出改进方案。</p> <p>主要参考文献：</p> <p>[1] K. Lee, The research-practice gap in user authentication, Ph.D. Thesis, Princeton University, 2022.</p> | 王琦 |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--------------|---------------------|-------------------|
| 用AIGC模型制作电影 | 运用最新的AIGC技术，探索制作电影 | 危学涛 |



| Title | Introduction | Supervisor |
|------------------|---|------------|
| 人类-机器文本检测任务 | <p>如何有效地区分真实人类创作的文本与大语言模型（机器）生成的文本？</p> <ul style="list-style-type: none">• 基于监督学习的文本分类方法是有效的，但是缺少通用性，需根据具体场景（文本类型、体裁、长度等）训练专门的分类器。• 零样本检测方法，如Fast-DetectGPT, FourierGPT等，具有通用性强、检测速度快、准确率高的优点。• 目前问题：中文数据集缺失，适用于中文文本的检测方法细节不明（例如，用何种推理模型，何种分词方法等等） <p>目标：创建中文文本检测的测评数据集；检验现有文本检测方法在该测试集上的表现；探索中文文本检测的技术细节。</p> | 徐炀 |
| 次级语言信号的语义分析 | <p>自然对话中存在大量次级语言信号，如停顿、语调变化、呼吸、非正式语言（back-channel）等。</p> <ul style="list-style-type: none">• 这些次级语言信号是否蕴含有可解释的语义？• 是否可以通过大语言模型（微调或其它方法）的隐藏表征层来解读它们的语义？ <p>目标：在对话数据集上微调一个较大型的语言模型（2b~7b），需根据特定数据集的标注方式向语言模型的词汇表增加表示次级语言信号的词符（token）；解读次级语言信号的隐藏表征。</p> | |
| 眼动指标与文本似然度的相关性研究 | <p>眼动指标是指阅读文本时，眼球在不同词语上的注视时长，以及从词到词之间的扫描路径。</p> <ul style="list-style-type: none">• 文本似然度指语句中每个词语的条件概率 $P(w_i w_{<i})$。• 以往研究表明眼动指标与文本似然度之间存在显著相关。 <p>目标：在中文眼动数据集上验证二者之间的相关性；寻找适合中文的文本似然度计算方法。</p> | |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--|--|--------------------|
| Unmixing in-the-wild multi-source mixtures | Realistic auditory environments usually contain multiple concurrent sound sources, which often dramatically degrade machine perception. This project aims at (1) developing deep learning based algorithms that can separate any target sound of interest (such as music, speech, animal sound, door closing etc.) from realistic in-the-wild real-recorded mixtures (such as the ones in the AudioSet); and (2) improving downstream tasks such as sound event detection based on the separation results. | Dr. Zhong-Qiu Wang |
| Unsupervised speech separation | Speech separation is a task aiming at separating multi-speaker mixtures to individual speaker signals. The current dominant approach is deep supervised learning based on simulated mixtures, but the trained models usually exhibit very limited generalizability to real-recorded mixtures. This project aims to improve the generalizability via unsupervised separation. | |
| Target sound extraction on hearables | We usually have to stay in crowded acoustic environments with multiple sound source signals such as annoying noises, unwanted speech, and music. To improve our well-being, this project aims at developing resource-efficient deep learning algorithms that can run on modern hearables, that can boost any sound source of interest or mute any undesired signals, and that can deliver spatial audio experiences for the wearers of the hearables. | |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|--|--------------------------|
| Speaker diarization and separation on conversational mixtures | In conversational scenarios (like in a meeting / dinner party with multiple people), we often want to diarize the conversations and figure out <i>who said what at when and where</i> . This project plans to leverage deep learning techniques and audio signal cues to improve the diarization performance on realistic conversational recordings, and in addotion leverage better diarization to improve speech separation. | Dr. Zhong-Qiu Wang |
| Audio-visual speech processing | Visual modality can help many audio processing tasks such as speech enhancement, separation, and recognition. This project aims to develop deep learning algorithms that can better leverage visual cues for speech processing. | |



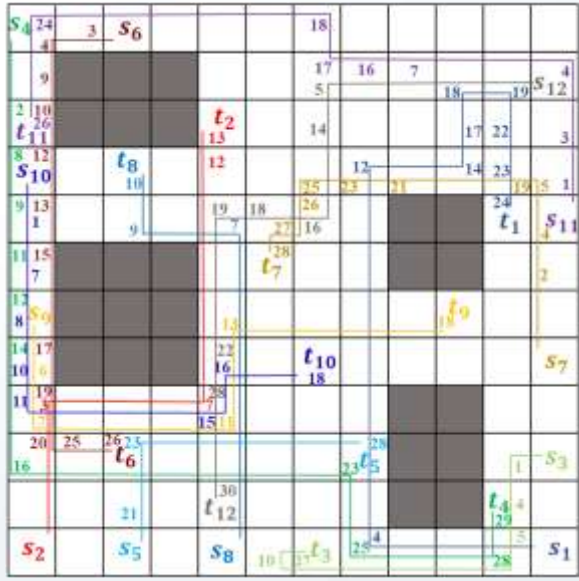
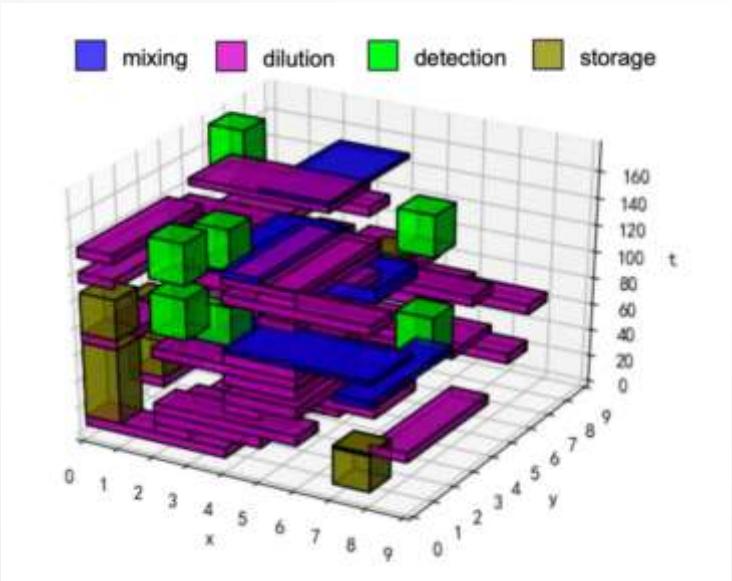
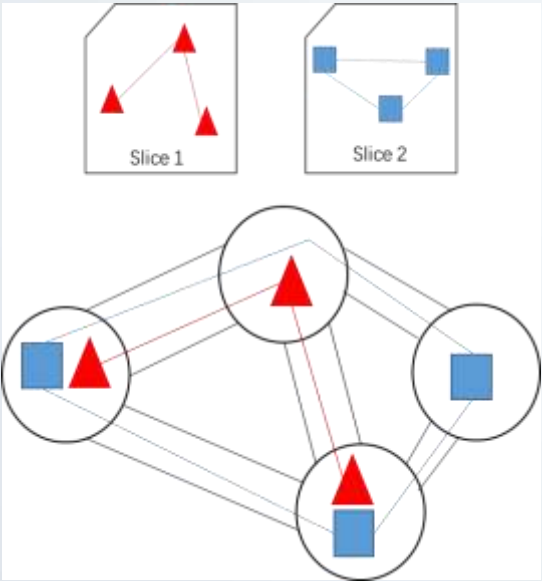
| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|-------------------|--|-------------------|
| 端到端加密服务的 多设备支持 | <p>端到端加密服务保证了用户之间的信息不被第三方获取，例如端到端通信软件保证了即使是服务提供商也无法窃取用户得通信内容。然而在现实应用中，每个用户可能拥有多个设备，如何在保证端到端安全的前提下进行多设备支持（如设备间数据的传输）是一个重要的问题。本课题主要是学习理解现行的多设备支持方案，分析其安全性，并进行开发和功能优化尝试。</p> <p>注：本课题涉及适合数理基础好的同学，开发部分主要是对算法理解的验证和对功能优化的测试。</p> | 陈杉 |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--------------------------|--|-------------------|
| <p>端到端加密服务的 密钥透明</p> | <p>端到端加密服务的安全性依赖于用户之间知道相互的身份标识（比如用户的公钥），而现实应用的协议大多要求用户去人工比对验证码（比如用户公钥生成的二维验证码），该操作比较笨重而且现实中也没有多少人去真正进行比对。针对此问题，人们提出了所谓的密钥透明，用来对用户的身份进行“自动”验证。本课题主要是学习理解现行的密钥透明方案，分析比较不同方案的安全性和效率，并进行开发测试。</p> <p>注：本课题涉及更高端的密码算法（比如零知识证明），适合数理基础好且对密码算法有浓厚兴趣的同学，开发部分主要是对算法理解的验证和算法分析的测试。</p> | <p>陈杉</p> |



| Title | Introduction | Supervisor |
|-------------|--|------------|
| 基于强化学习的组合优化 | 求解组合优化问题通常采用启发式算法或元启发式算法，然而这些启发式策略是由领域专家设计的，非常依赖人的经验；此外，由于组合问题的复杂度（多为NP难），这类方法一般来说不是最优的，或者求解过程非常耗时。强化学习（Reinforcement Learning）被认为是一种求解组合优化问题的非常好的选择，通过使用监督或自我监督的方式训练智能体，从而自动实现启发式策略。本项目拟探索强化学习在实际复杂场景下（多目标/多约束/多任务）的组合优化问题中的应用，如映射、调度、布局、路由等。 | 袁博 |





| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|---|-------------------|
| Research on accelerating computations on GPUs | <p>Background: GPUs, originally developed for rendering graphics, have evolved into powerful parallel processors capable of handling thousands of simultaneous threads. This makes them ideal for tasks that require significant computational power, such as scientific simulations, data analysis, and machine learning. The significance of GPU acceleration lies in its ability to drastically reduce computation times. By parallelizing tasks, GPUs can perform many calculations concurrently, leading to faster data processing and real-time performance in applications where speed is critical.</p> <p>Goal: In this project, we will start from some real-world scenarios and explore different techniques for optimizing the computations in these real-world scenarios to fully exploit the parallel nature of GPU.</p> | Zhuozhao Li |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---|---|-------------------|
| Performance interference study of co-located applications | <p>Cloud providers often co-locate latency-critical and batch applications together on the same physical machines to increase resource utilization.</p> <p>In this project, we will explore the impact of interference for applications running on co-located machines. We will study applications with different characteristics and investigate the performance interference of various application combinations.</p> | Zhuozhao Li |
| Research on optimizing the performance of AI-driven HPC workflows | <p>Integrating AI systems into traditional high-performance computing workflows has demonstrably enabled highly accurate modeling and holds significant promise for accelerating scientific discovery. In this project, we will understand the definition and primary performance challenges of AI-driven HPC workflows. We will explore various frameworks to ease the coupling of AI and HPC components in a single workflow and research on optimizing the performance of AI-driven HPC workflows.</p> | Zhuozhao Li |

一、背景资料

科学与工程计算中心负责管理和运维南科大高性能计算平台1100多台服务器。由于服务器数量众多，构建集群软硬件监控和告警系统对于保障集群的稳定运行、及时发现和处理故障、提高系统性能和预防潜在问题至关重要，同时也能够提高运维工程师的工作效率。目前，已搭建了一套基于Prometheus的监控和告警系统，需要进一步完善系统功能。

二、课程实践的目标

- 1) 中心将提供现有系统源代码以及测试服务器资源。
- 2) 与计算中心工程师协同完成需求分析，结合计算中心的实际需求，完成开发系统。
- 3) 开发高性能集群监控和告警系统，完成：
 - a、监控数据的采集（包括硬件告警和操作系统运行数据）；
 - b、数据存储和分析；
 - c、告警规则设置；
 - d、系统用户管理和告警处理功能；
 - e、提供简洁美观、使用友好的用户界面。

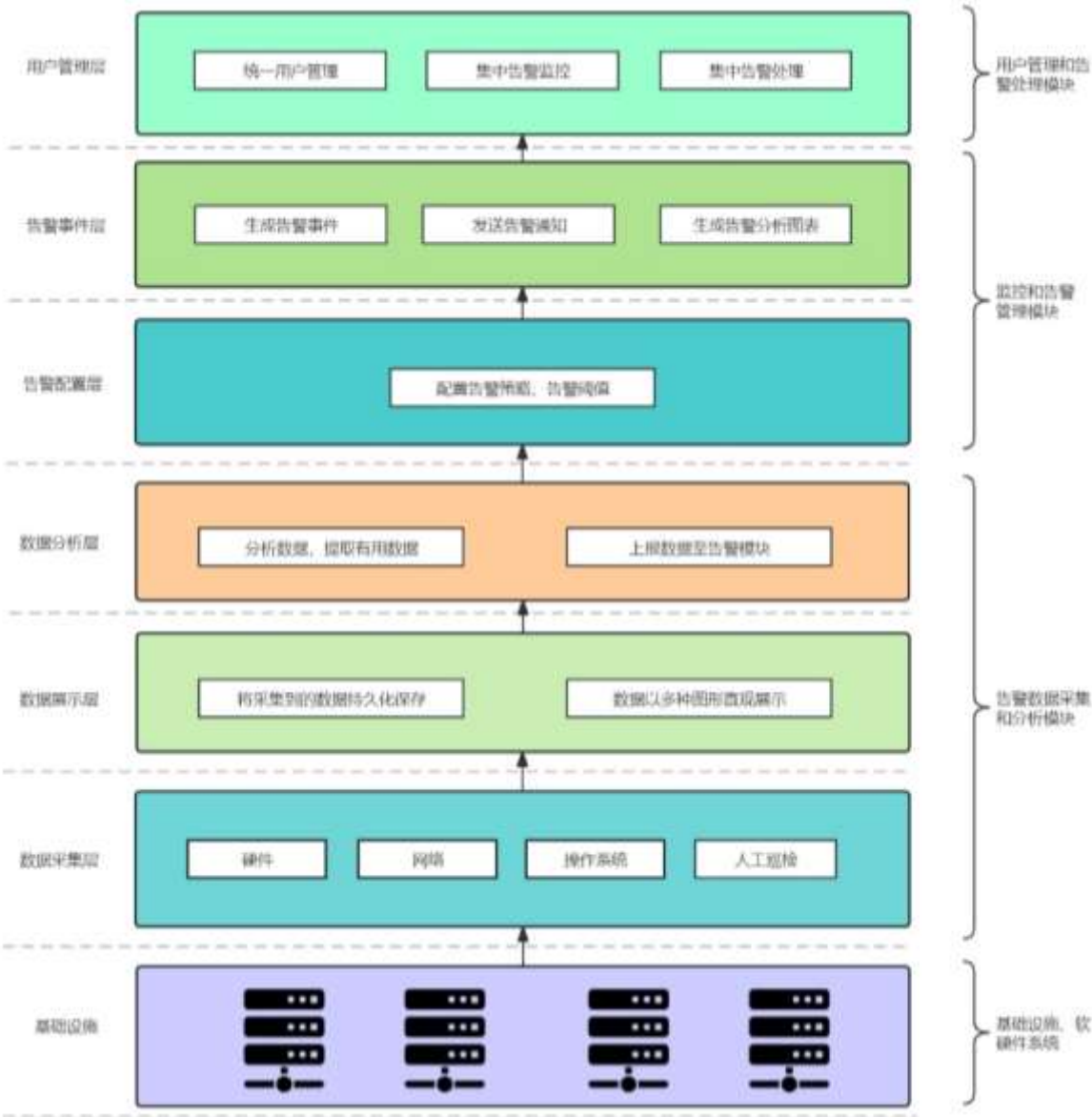


图1、高性能集群监控和告警系统架构图



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|--------------|---|----------------------|
| 超算门户与管理系统开发 | <p>背景资料</p> <p>南科大计算平台是学校公共科研计算平台，平台目前拥有1000多个计算节点，4万多计算核心，50多张GPU卡，每月用户计算作业量超过10万。目前需要开发方便用户使用和运维人员部署且功能齐全的超算门户与管理系统。</p> <p>项目内容</p> <ul style="list-style-type: none">• 搭建异构化高性能计算平台；• 现代化富有亲和力前端UI设计；• 门户功能设计与实现：集成 web shell、文件管理、作业提交、可视化应用等一站式服务；• 管理功能设计与实现：用户账户管理；计算资源计费收费（作业）；账户、用户限额等• 其它要求：采用现代技术栈和部署方案，方便部署、维护、自定义和扩展。 | 李卓钊、南科大计算中心的范靖、杨期垚老师 |



| Title | Introduction | Supervisor |
|---|--|------------|
| Federated Learning for Unsupervised Health Analytics | This project focuses on leveraging federated learning to enable unsupervised health data analytics across multiple consumer-grade wearable sensors (e.g., smartwatches) without compromising patient privacy. Federated learning allows models to be trained collaboratively on data distributed across various locations, ensuring that sensitive information remains localized. Students will work on developing algorithms that can identify patterns and anomalies in health data in an unsupervised manner, aiming to improve early diagnosis and personalized treatment plans while adhering to stringent privacy regulations. | 郑天越 |
| Intelligent Earphone Binaural Audio Sensing | This project aims to utilize the dual-microphone arrays in modern Bluetooth earphones for sensing various user actions, such as head movements and hand gestures. By capturing and analyzing binaural audio signals, this project seeks to extract meaningful patterns using advanced signal processing and machine learning techniques. Potential applications include enhanced user interfaces, immersive virtual reality experiences, and improved health monitoring systems. This project will provide students with hands-on experience in cutting-edge audio processing technologies and help them develop vital skills in machine learning, signal processing, and the development of innovative wearable technologies. | |
| Fluid Antenna Design for Integrated Sensing and Communication Systems | Students undertaking this project will delve into the innovative realm of fluid antennas, which have the ability to dynamically change their configurations to optimize performance. The project will involve designing fluid antennas that can seamlessly integrate sensing and communication functionalities, offering enhanced adaptability and efficiency in various applications such as IoT, smart cities, and autonomous vehicles. The goal is to develop prototypes that demonstrate the practical benefits of fluid antennas in real-world scenarios. | |
| Adapting Foundation Models for RF Sensing Tasks | This project aims to adapt large-scale pre-trained models, known as foundation models, for specialized tasks in the domain of radio-frequency (RF) sensing. Foundation models, which have shown remarkable capabilities in fields like natural language processing and computer vision, hold promise for enhancing RF sensing applications such as object detection, activity recognition, and environmental monitoring. Students will focus on modifying and fine-tuning these models to effectively process and interpret RF signals, contributing to advancements in wireless communication and sensor networks. | |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|---------------------|---|-------------------|
| 金融市场仿真器参数校准的预训练加速方法 | 金融市场仿真在使用前需要对历史数据进行校准，这可以被建模为统计推断或者参数优化问题，由于灰盒属性，通常需要迭代求解（迭代采样或搜索），这需要大量次数的运行仿真器，极为耗时。本项目拟研究更加高效的校准方法，通过构建预训练模型来预测不同仿真器参数的表现，进而减少真实仿真器运行次数，最终实现校准加速的目的。 | 杨鹏 |
| 金融市场模拟的多场景交互展示界面研发 | 调研纳斯达克金融仿真器的几个典型应用场景，理解市场行情数据的用途与特性，设计并研发金融仿真器可交互式展示界面。 | |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|-----------------------|--|-------------------|
| LLMs for CS education | Applying LLMs to aid various teaching and learning activities, and to boost productivity | 陶伊达 |

张煜群组创新实践项目



| Title | Introduction | Supervisor |
|------------------|---|------------|
| 基于大模型的代码生成技术研究 | 研究基于大模型的代码生成技术。课题组已经和快手公司合作，产生了在工业界较为领先的代码生成技术。在该技术基础上，将会继续推进研究进程，取得更优的代码生成效果。从事该研究的同学有机会获得充分的算力资源，并在暑期进入该公司进行实习。 | 张煜群 |
| 基于大模型的反编译技术研究 | 课题组于2024年3月发布全世界第一个开源反编译大模型。一经发布即吸引大量关注。迄今已经在GitHub上收获超过2.8K星，并曾在Trending榜上和Open Sora并列前三。在性能上也轻松取得比GPT4显著优越的效果。在此基础上，将会和字节跳动公司合作，开发面向实际应用的强力反编译大模型。从事该研究的同学有机会获得充分的算力资源，并在暑期进入该公司进行实习。 | |
| 基于大模型的安全漏洞挖掘技术研究 | 课题组和华为合作，利用大模型强大的学习能力，从已知系统的漏洞发布中进行学习，对目标系统的安全漏洞进行快速高效挖掘。从事该研究的同学有机会获得充分的算力资源，并在暑期进入该公司进行实习。 | |
| 软件安全基础设施技术研究 | 课题组和华为合作，基于课题组长期以来在模糊测试等软件安全技术研究的积累，向底层出发，研究促进整体技术发展的基础设施优化技术，例如如何更高效的对程序插桩等等，从而彻底的提升模糊测试等软件安全技术的实用价值。从事该研究的同学有机会在暑期进入该公司进行实习。 | |



| Title | Introduction | Supervisor |
|-------------------------------|---|------------|
| 面向大语言模型的生成内容优化 | <p>当前，大语言模型(LLMs)凭借着其出色的文本生成能力已经成为一项开创性技术，但其生成内容的准确性和适应性无法得到保障。尽管借助预训练和模型微调可以缓解该问题，但其需要消耗大量的计算资源、花费大量的训练时间。本项目将基于提示工程和干预方法，复现以下论文的实验，以在不改变模型参数的前提下提升大语言模型输出内容的质量。</p> <p>参考文献： Self-consistency improves chain of thought reasoning in language models ; Large Language Models Cannot Self-Correct Reasoning Yet</p> | 唐茗 |
| 边缘智能设备在应对复杂工况时面临的挑战及潜在解决方案的探究 | <p>边缘智能（ Edge Intelligence ）是一种结合边缘计算和人工智能技术的计算范式。其核心思想是在靠近数据源的边缘设备上进行数据采集、数据处理和智能分析，而不是依赖于远程的集中式数据中心或云计算平台。通过在设备本地进行数据处理，边缘智能能够显著减少数据传输的延迟和带宽需求，提高实时响应能力和隐私保护水平。但是，边缘设备的性能通常会受到所处环境的影响，本项目以自动驾驶汽车为例来探究边缘设备在面临各种恶劣工况时的挑战以及潜在的应对方案。</p> <p>参考文献： Perception and sensing for autonomous vehicles under adverse weather conditions: A survey</p> | |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|-------------------|--|-------------------|
| 基于Rust的安全操作系统 | 本项目基于课题组与蚂蚁集团合作研发的Rust安全操作系统Asterinas和RIPOS，使用Rust编程语言，构建安全操作系统的功能模块。（要求：学生需要有能力学习Rust编程语言，对学习和搭建操作系统有浓厚兴趣。） | 张殷乾 |
| AI和大模型的系统安全架构 | 本项目将着眼于现有人工智能系统架构，特别是大模型系统中的安全问题展开研究，目标是构建下一代安全AI系统软件。（要求：学生需要有C/C++语言编程能力，对学习和搭建系统软件有浓厚兴趣。） | 张殷乾 |
| 机密计算的web3和隐私计算研究 | 本项目将基于机密计算和可信执行技术打造web3和隐私计算的应用。（要求：具有编程能力和独立探索能力。对web3、区块链、隐私计算技术感兴趣。） | 张殷乾，牛健宇 |
| 强化学习的区块链性能安全和性能增强 | 本项目将使用强化学习技术去增强区块链平台，尤其是分布式共识算法的性能和安全，分析现有平台激励问题和分布式动态调度。（要求：具有一定编程能力和独立探索能力。对分布式系统、区块链或者强化学习感兴趣。） | 张殷乾，牛健宇 |



| <i>Title</i> | <i>Introduction</i> | <i>Supervisor</i> |
|----------------------|---|-------------------|
| 基于大型语言模型的优化算法发现/自动设计 | 本项目旨在探索利用大型语言模型来辅助算法发现和自动设计的可能性，探索的方向包括算法结构生成、启发式函数生成、算法性能优化建议生成等，采用的技术路线包括但不限于提示工程、RAG等。 | 刘晟材 |