



云计算与数据科学 研究所

负责人：孙永奇

<http://ccd.bjtu.edu.cn>



研究所概况

- **中国工程院院士、顾问教授1人；**
教授3人，副教授2人，讲师1人；
在读硕士、博士研究生50余名。
- **研究云计算与云原生、分布式人工智能、数据挖掘与知识图谱、深度学习与强化学习等理论及应用技术。**
- **负责北京市大数据人才培养示范基地的建设。**
- **负责BJTU-AWS云创学院的建设。**
- **负责《数据结构》、《大数据技术》、《并行与分布式计算》课程建设。**

云计算与分布式计算

高性能计算的范围很广，包括：

- 并行计算（大型机，MPI）
- 多核计算（OpenMP）
- 分布式计算（**云计算**）
- 基于GPU的并行计算（CUDA架构）
- 众核计算（混合架构， OpenMP+MPI）



Google云计算

Google的云计算梦想

应用向互联网迁移

数据向互联网迁移

计算能力向互联网迁移

存储空间向互联网迁移



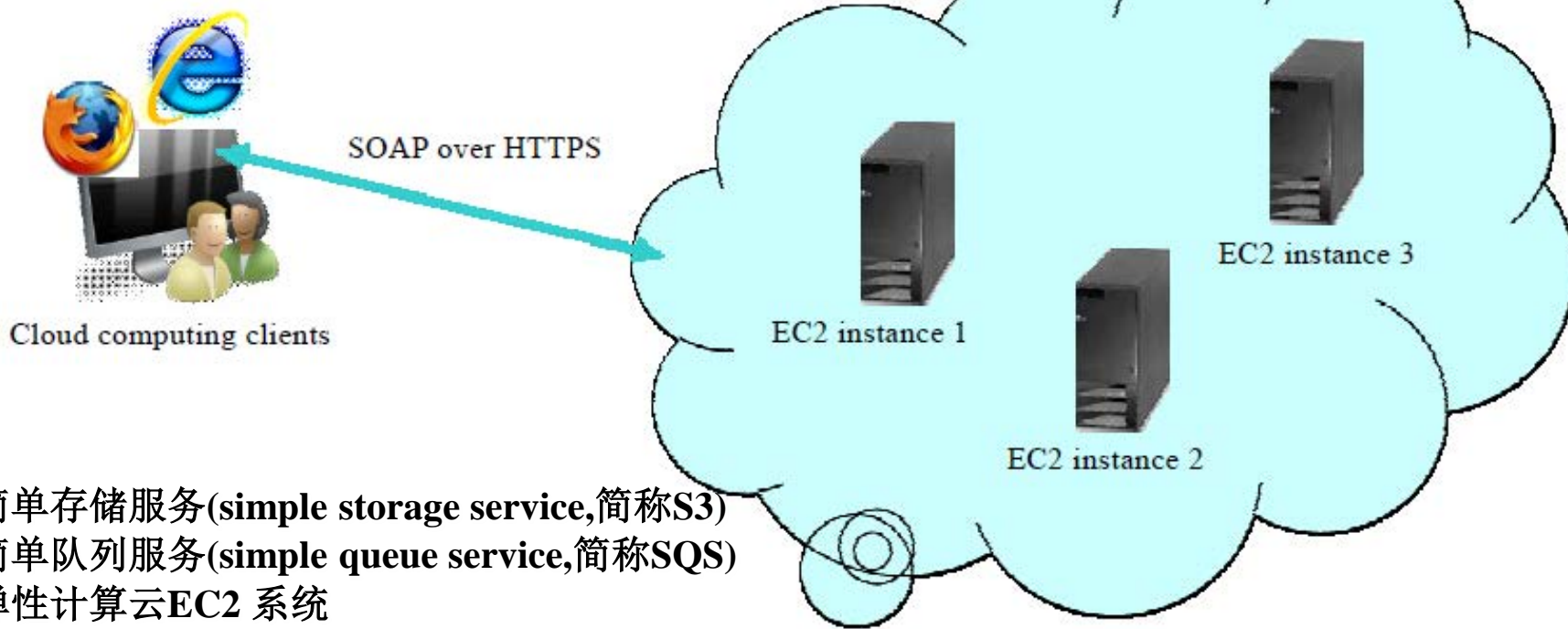
Chrome OS

“浏览器=操作系统”

亚马逊云计算AWS

关键功能

Amazon elastic computing cloud (EC2)

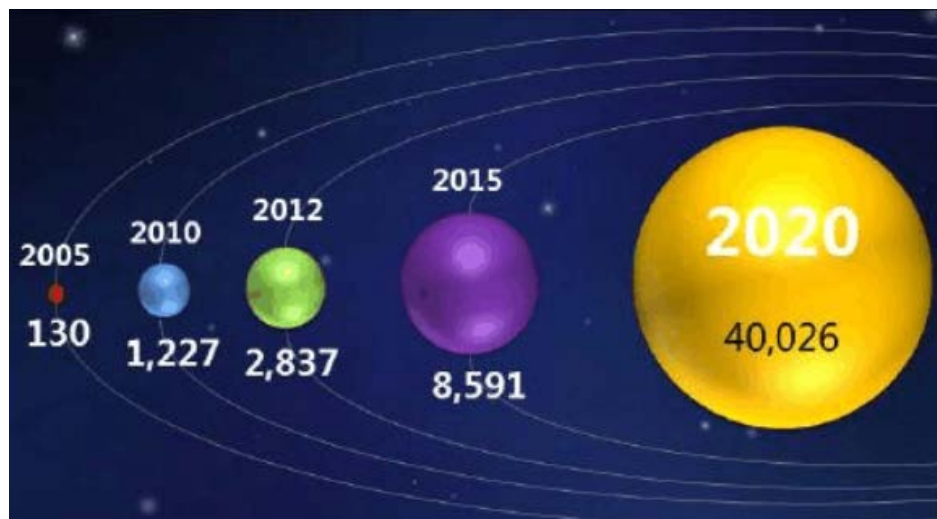


简单存储服务(simple storage service,简称S3)
简单队列服务(simple queue service,简称SQS)
弹性计算云EC2 系统

数据的发展

数据正以每年超过 50% 的速度爆发式增长。IDC 最新的研究报告显示，到 2020 年，全球的数据资料存储量将超过 40ZB。

大数据的特点 (4V)：量大 (Volume)、种类多 (Variety)、产生和处理速度快 (Velocity)、价值密度低 (Value)





股票交易



实时车载信息



卫星勘测



超市交易



时间序列数据

Internet通信



Present of Big Data

Too big to handle



大数据与云计算

- 大数据为体
- 云计算为术



一、研究所主要成员介绍

赵春江 院士

- 研究所学术带头人，北京交通大学顾问教授
- 中国工程院院士
- 研究方向：
人工智能
数据挖掘与数据库技术
智能化专家系统
模型与决策支持系统



一、研究所主要成员介绍

王志海 教授

- 博导，北京市教学名师

- 研究方向：
数据挖掘与商务智能
数据库与数据仓库
机器学习与计算智能



➤ zhhwang@bjtu.edu.cn

孙永奇 教授

- 博导，研究所所长

- 研究方向：
大数据与云计算
人工智能
图神经网络
联邦学习



➤ yqsun@bjtu.edu.cn

一、研究所主要成员介绍

徐保民 教授

- 博导
- 研究方向：
分布式计算、云计算
数据库
区块链技术
高级Web开发技术



➤ bmxu@bjtu.edu.cn

原继东 副教授

- 博导，研究所副所长
- 北京市优秀人才（青年骨干个人）
- **20-23**连续四届毕业生“**我最敬爱的老师**”
- 研究方向：
时间序列数据挖掘
知识图谱
时间序列模型对抗攻击
深度学习、强化学习



➤ yuanjd@bjtu.edu.cn

一、研究所主要成员介绍

徐薇 副教授

- 博士，硕导
- 研究方向：
数据挖掘
机器学习
数据库



➤ wxu@bjtu.edu.cn

刘海洋 讲师

- 博士，硕导
- 研究方向：
时间序列数据挖掘
多标记学习
图数据挖掘
联邦学习



➤ haiyangliu@bjtu.edu.cn

二、主要科研项目

1、分布式人工智能、联邦学习

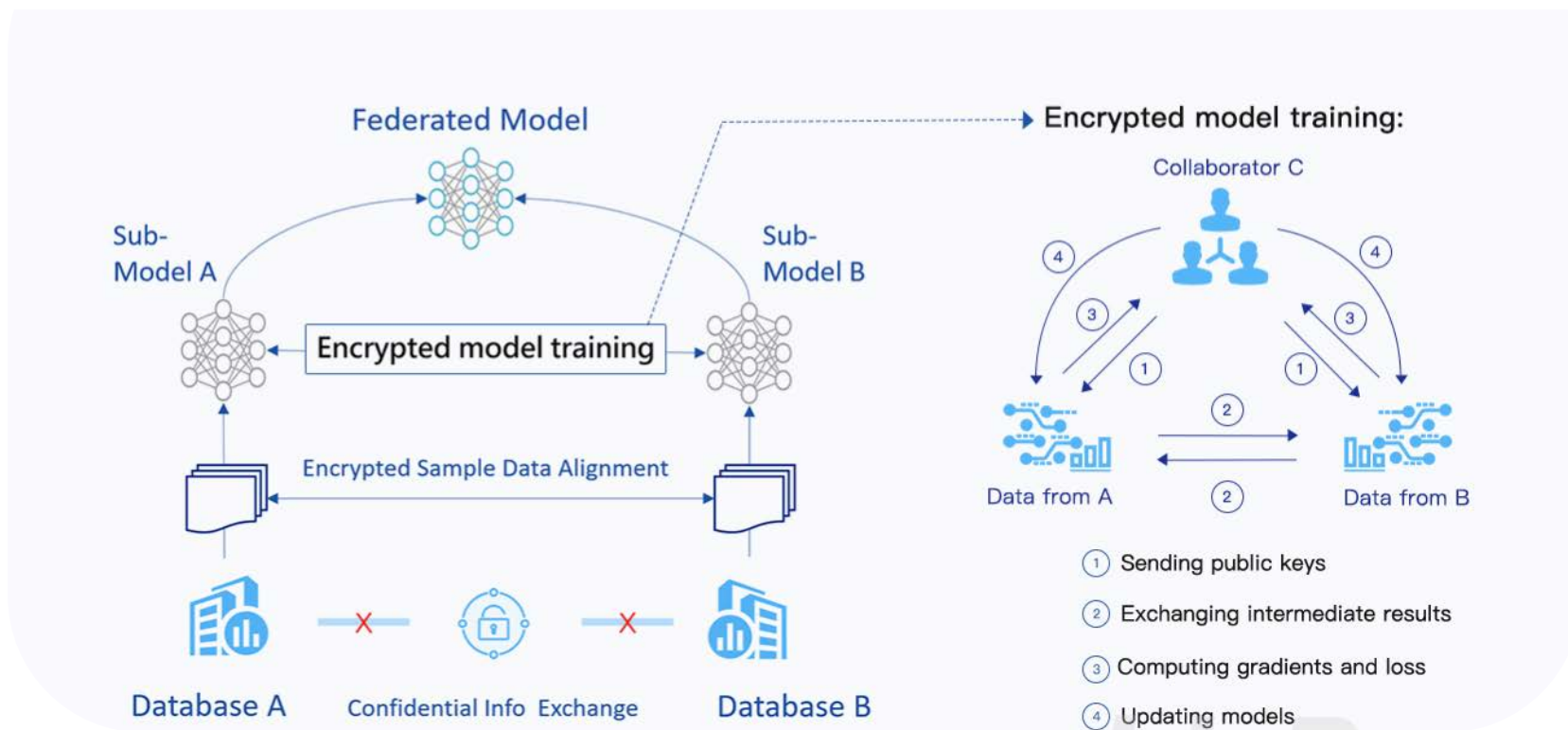
(孙永奇、刘海洋、原继东)

- 个性化联邦学习
- 联邦推荐系统
- 联邦学习节点大规模扩展技术
- 自适应、高效的隐私保护技术
- 落地平台：PaddlePaddle（百度）、MindSpore（华为）、Tensorflow、PyTorch



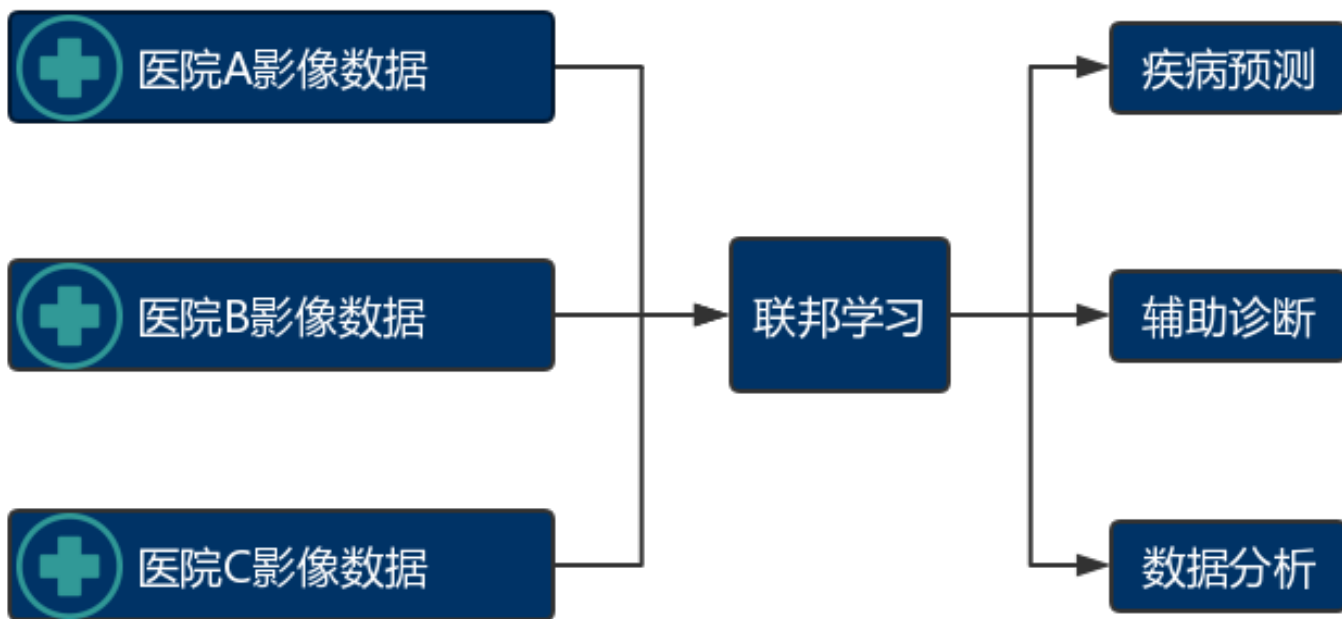
➤ 联邦学习系统架构

- 联邦学习：分布式计算+深度学习/机器学习，分横向、纵向、迁移三种
- 解决数据孤岛问题、保护用户的数据隐私
- 下图是两个数据拥有方的场景，可扩展至多个数据拥有方



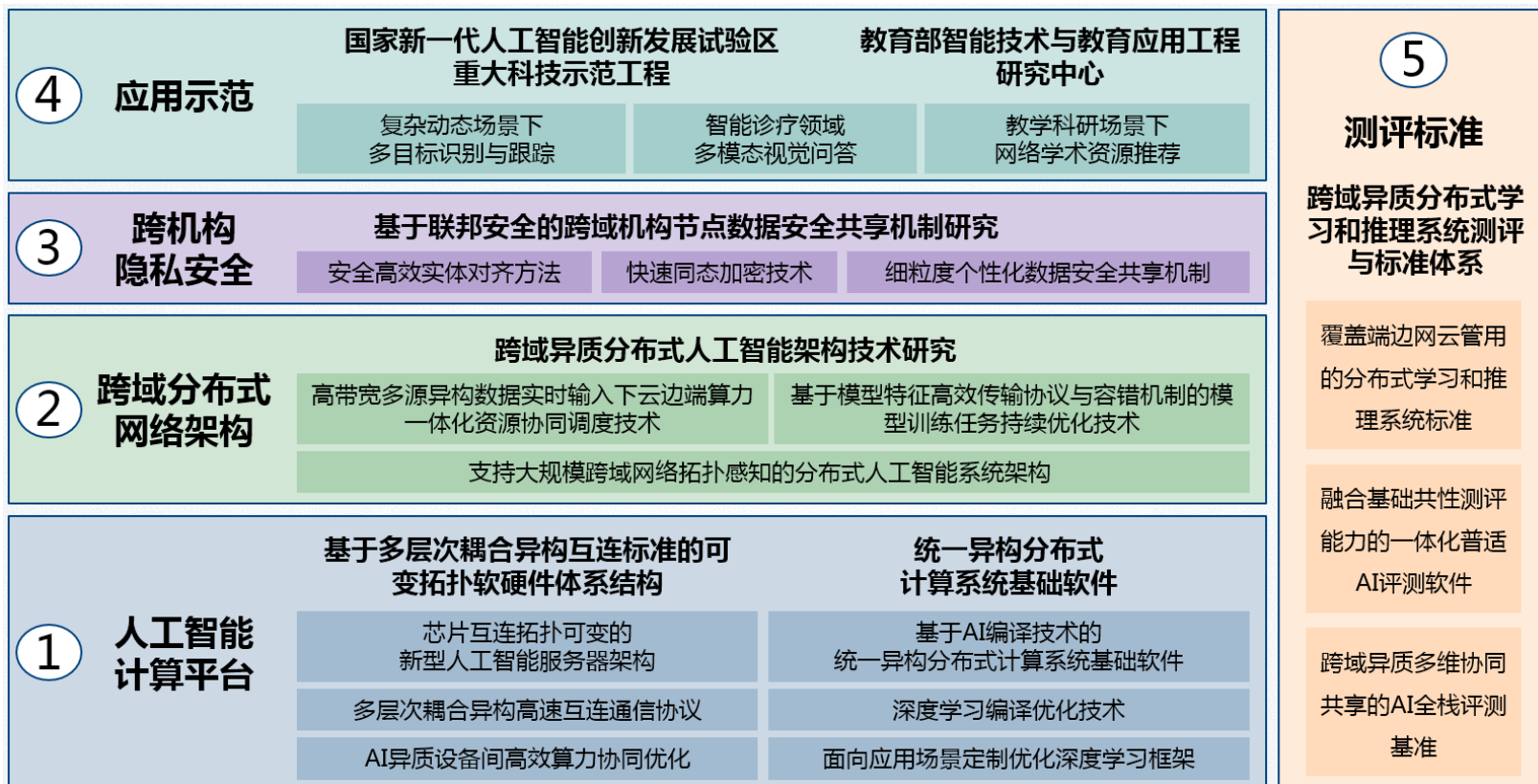
➤ 应用领域之一：图像识别

联邦学习在在医疗领域的医学图像识别中应用，可以解决医学数据孤岛问题，提升医学机构的人工智能应用。



➤ 跨域异质分布式学习和推理系统（与浪潮、百度合作）

■ 科技创新2030 “新一代人工智能” 重大项目，研究跨域异质分布式学习和推理系统面临的大规模跨域网络复杂化、跨机构数据隐私保护、AI异质设备多样化等重大科学难题。



2、大规模时间序列分析与预测

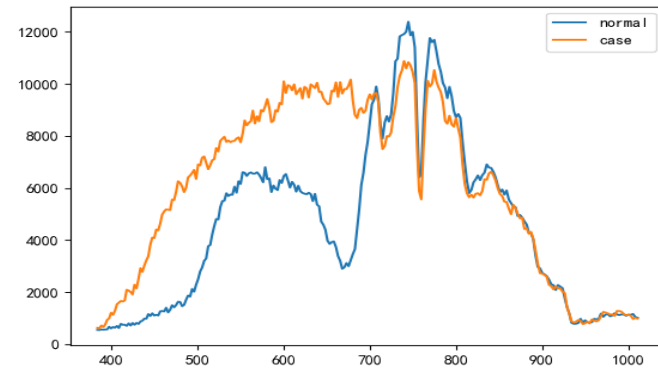
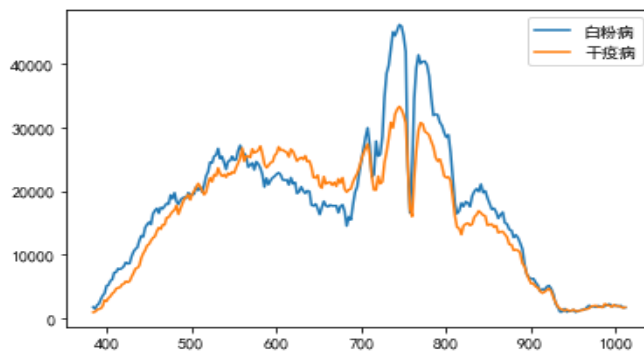
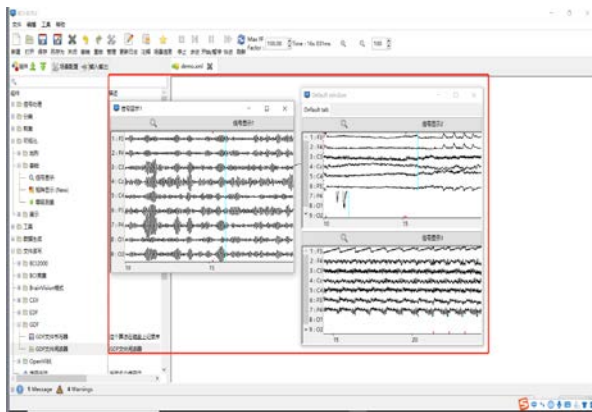
(原继东、刘海洋、孙永奇、王志海)

- 时间序列表示学习
- 时间序列辨别性模式发现
- 深度学习模型的对攻击与防御
- 时间序列分类与预测
- 数据流概念漂移检测
- 序列推荐与因果推荐系统



2、大规模时间序列分析与预测

- **应用场景：**基于高光谱时间序列的病虫害检测、可穿戴数据的快速识别与分类、用户特征分析、手写汉字识别、在线学习行为预测与分析以及交通流量、股票趋势、设备故障的预测等。



➤ 掌子面围岩条件实时智能感知理论-国家重点研发项目

- 挑战：基于历史工程建立的机器学习模型不适用于新建工程；新建工程数据量少，深度学习模型效果差
- 对策：基于**TBM破岩机理**的数据转换关系 + 数据**迁移学习**方法
- 意义：有效**利用历史工程的数据知识**指导新建工程的TBM施工



3、分布式对抗博弈与多智能体仿真技术

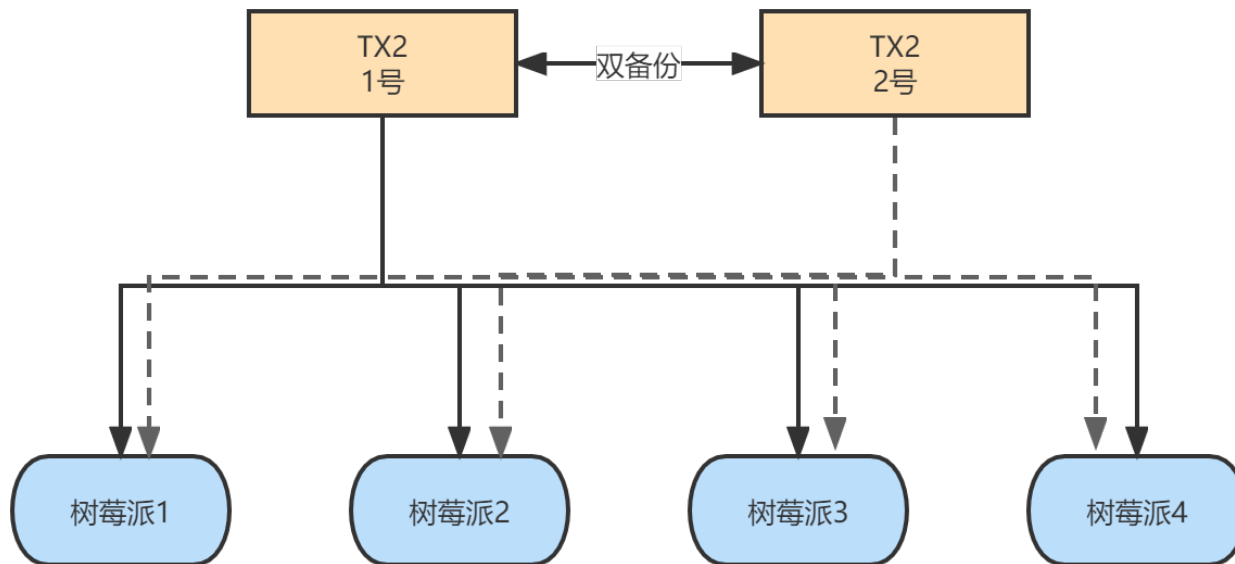
(刘海洋、原继东、孙永奇、徐保民)

- 利用Unity3D仿真生成能够克服环境和地域限制的飞机、车辆和轮船等的红外图像，有效降低研发成本。



3、分布式对抗博弈与多智能体仿真技术

- 将深度强化学习算法构建为微服务，基于分布式硬件资源的虚拟化完成统一调度与管理。



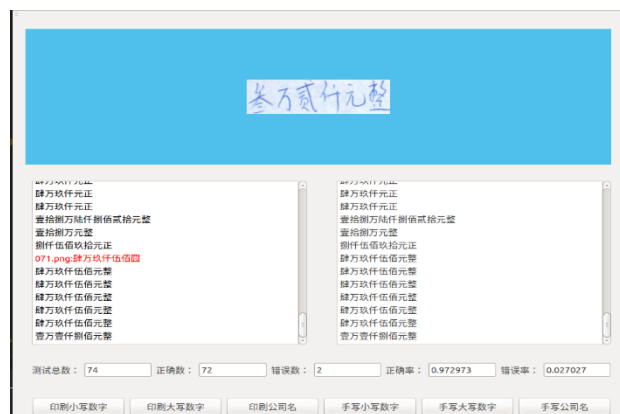
3、分布式对抗博弈与多智能体仿真技术

- 研究数字孪生与并行仿真技术，利用多智能体控制技术保证仿真环境真实性



4、其他项目

- 基于深度学习的**图像文本**、**手写票据内容识别**（孙永奇等）
- 基于文本挖掘的**舆情分析系统**（王志海、原继东等）
- 轨道交通车辆中心作业及设备**运维一体化平台**（徐保民等）



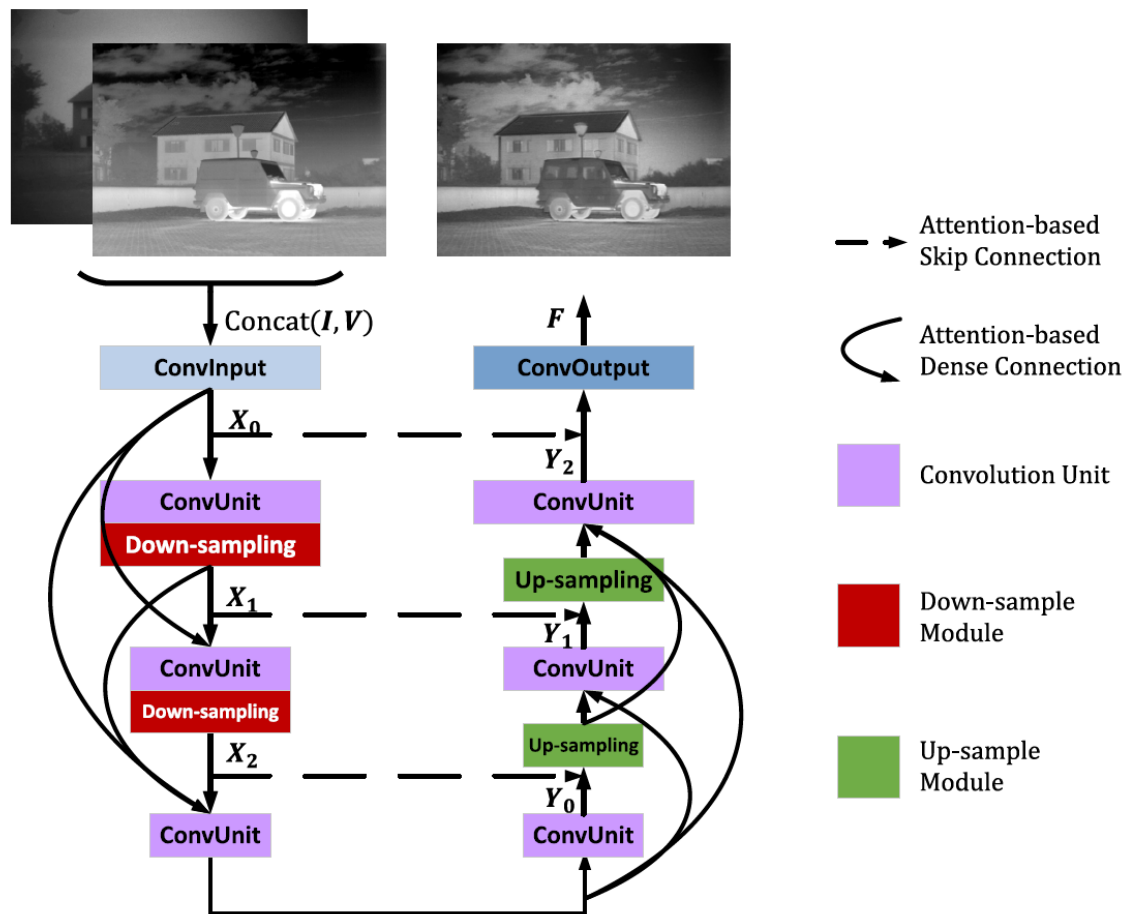
➤ PIS智慧运维平台项目

- 面向城市轨道交通的智能运维平台，利用智能化、信息化、可视化和大数据等手段，实时监控设备的运行状态、采集和下载PIS系统的运行数据。（徐保民、徐薇等）

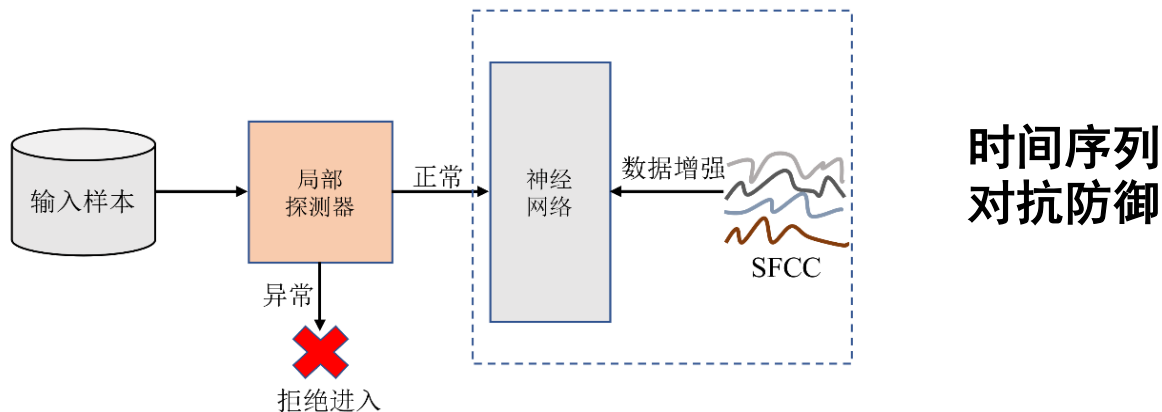
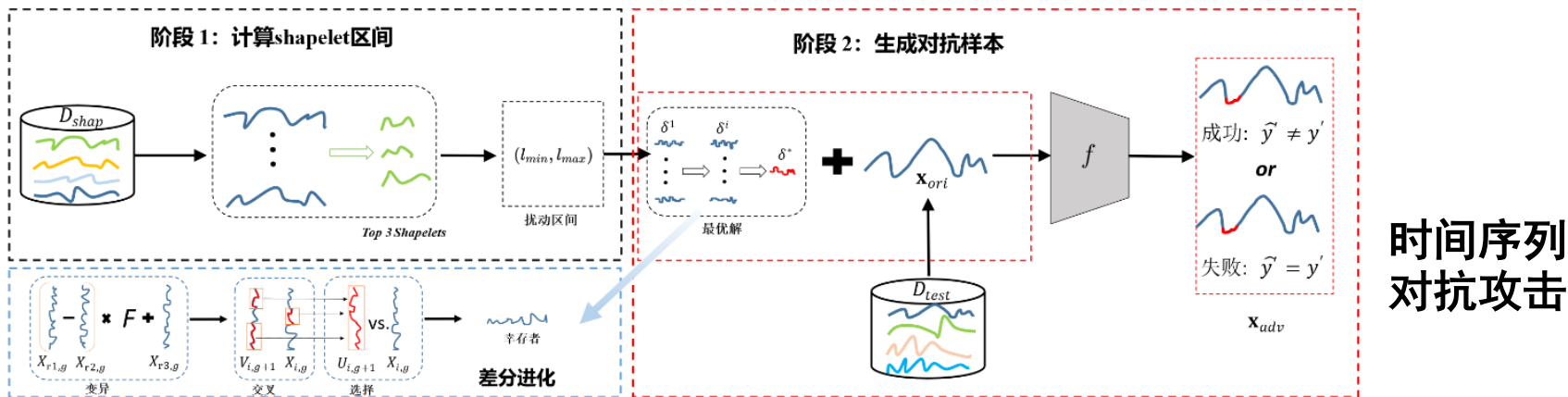


三、近年来的优秀硕士生成果

1. 程辰 (基于风格损失的可见光和红外图像融合模型, 被评为**2023届校优秀毕业生**)



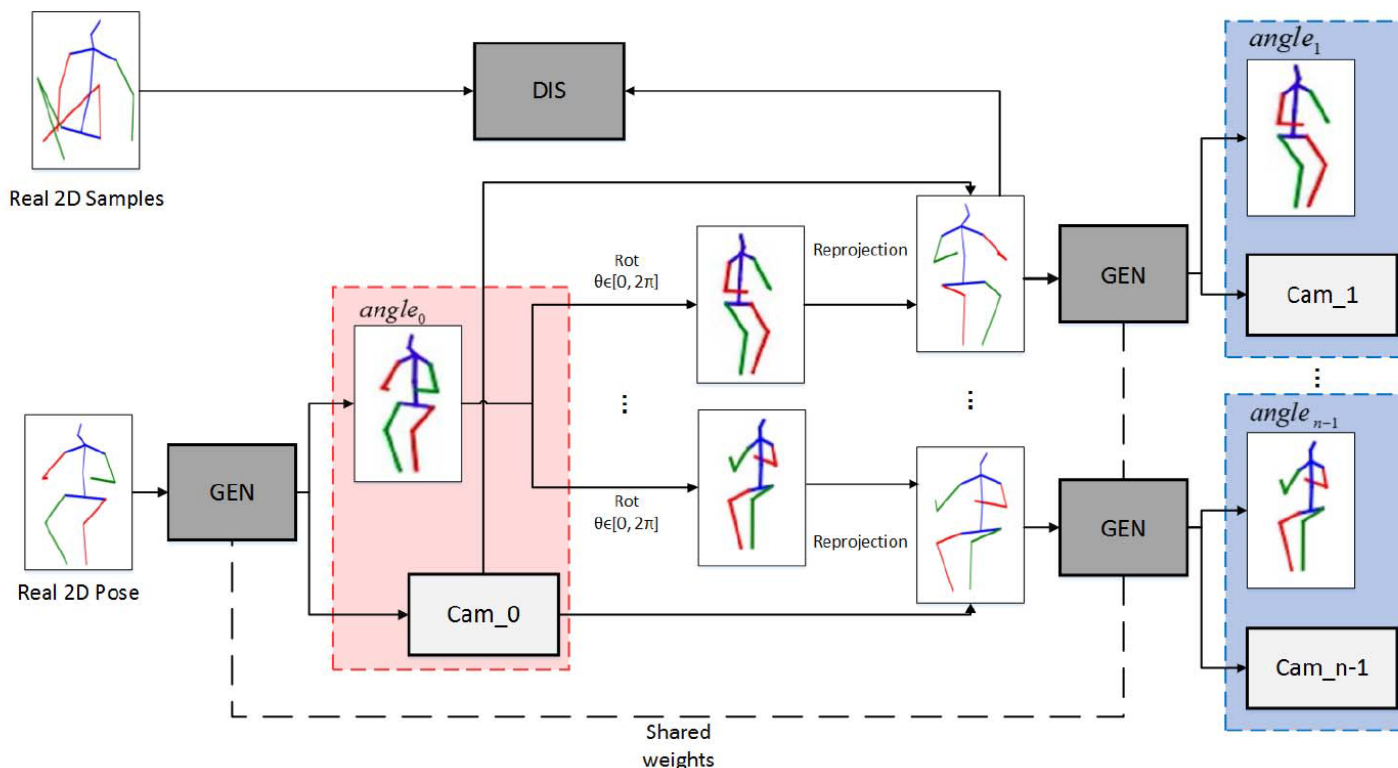
2. 杨文博 (基于局部扰动的时间序列对抗攻击及防御研究 2022届)



Wenbo Yang, Jidong Yuan et al. TSadv: Black-box Adversarial Attack on Time Series with Local Perturbations[J]. Engineering Applications of Artificial Intelligence. 2022, 114: 105218. 2022.09 (SCI - An2期刊)

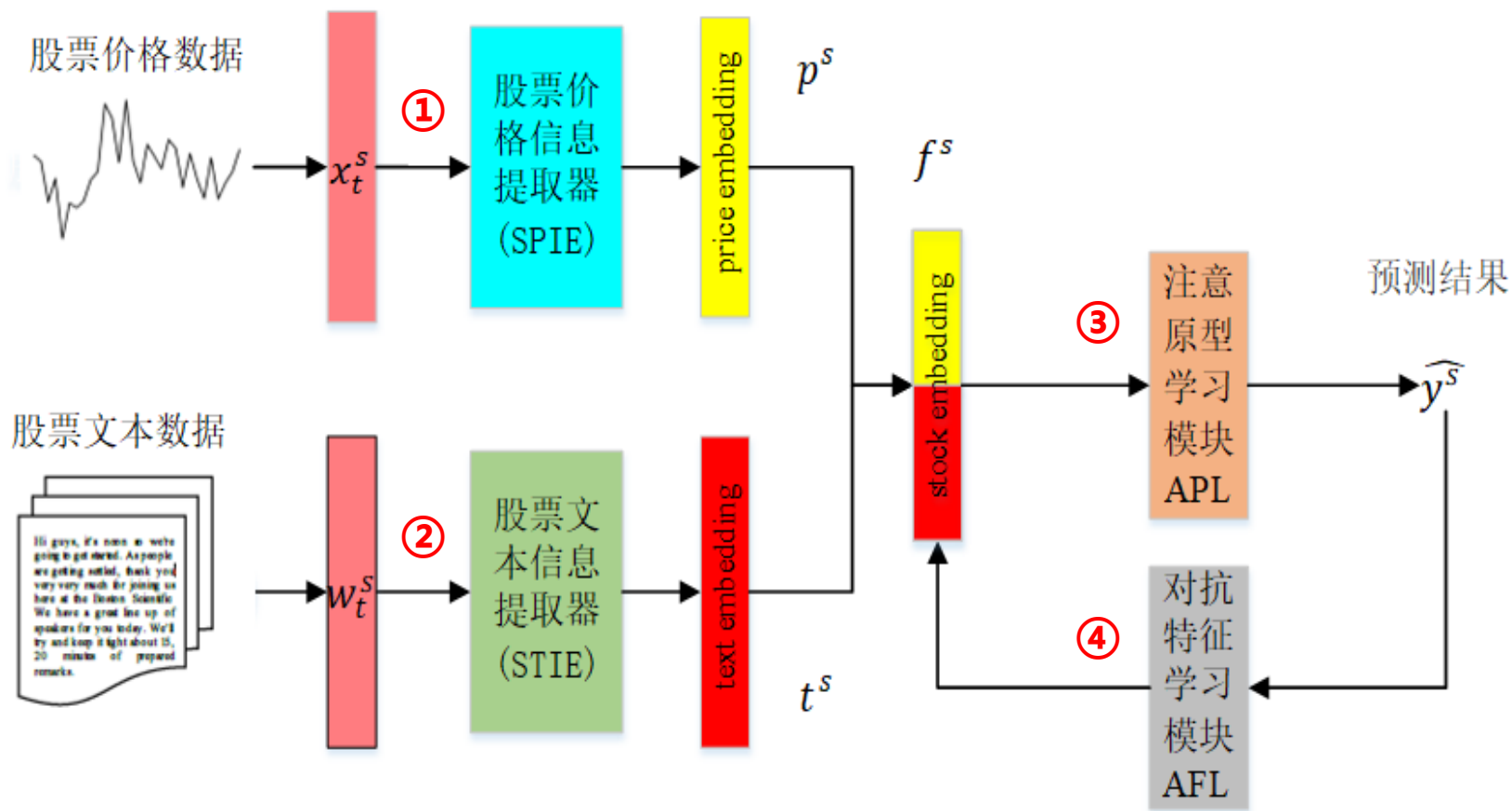
Wenbo Yang, Jidong Yuan et al. SFCC: Data Augmentation with Stratified Fourier Coefficients Combination for Time Series Classification[J]. Neural Processing Letters. 2022. 07 (SCI - An3期刊)

3. 邓祎程 (基于单视角多角度一致性的3D人体姿态估计模型, 2022届)

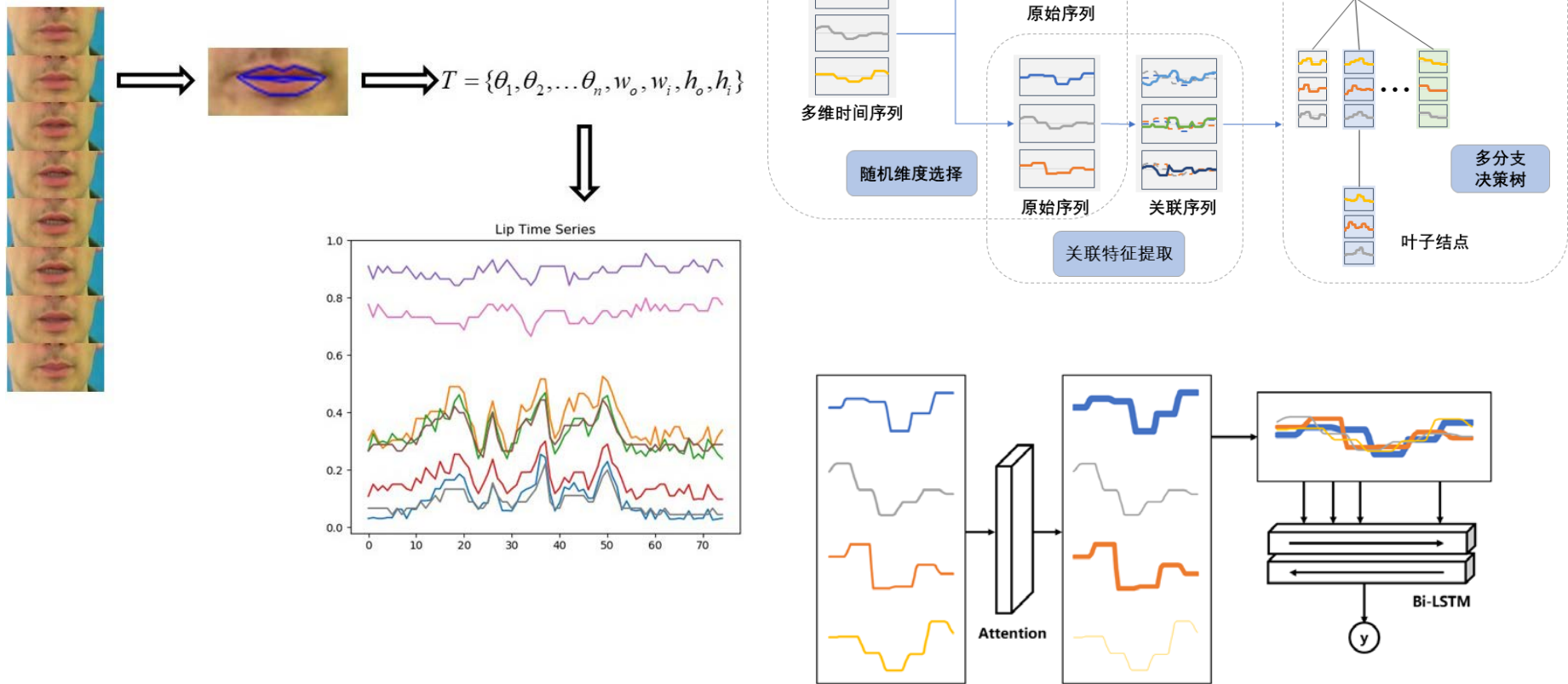


Yicheng Deng, Cheng Sun, Yongqi Sun*, Jiahui Zhu. SVMAC: Unsupervised 3D human pose estimation from a single image with single-view-multi-angle consistency. The 9th International Conference on 3D Vision, 2021.

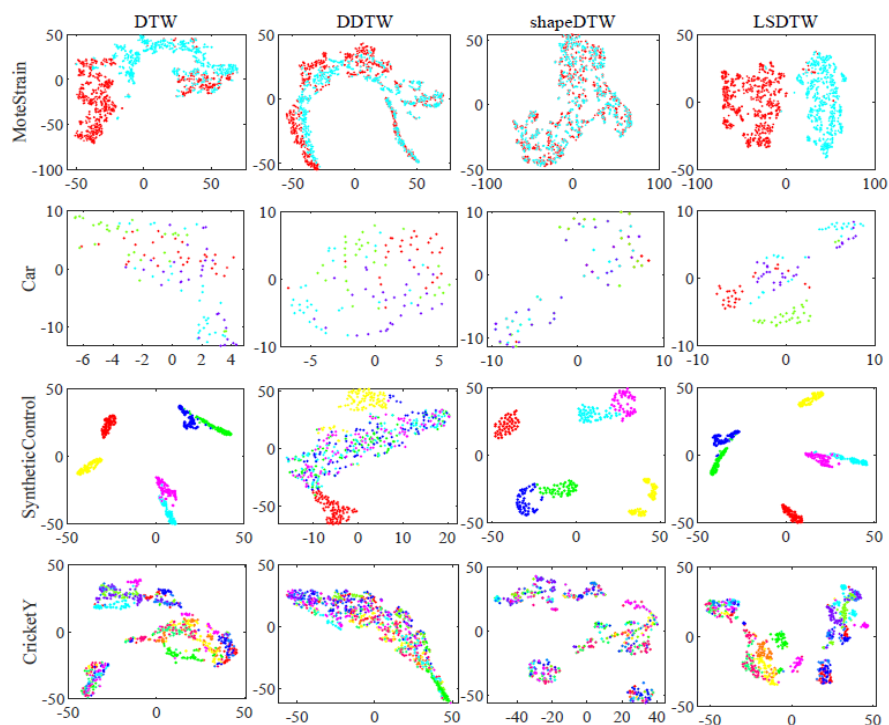
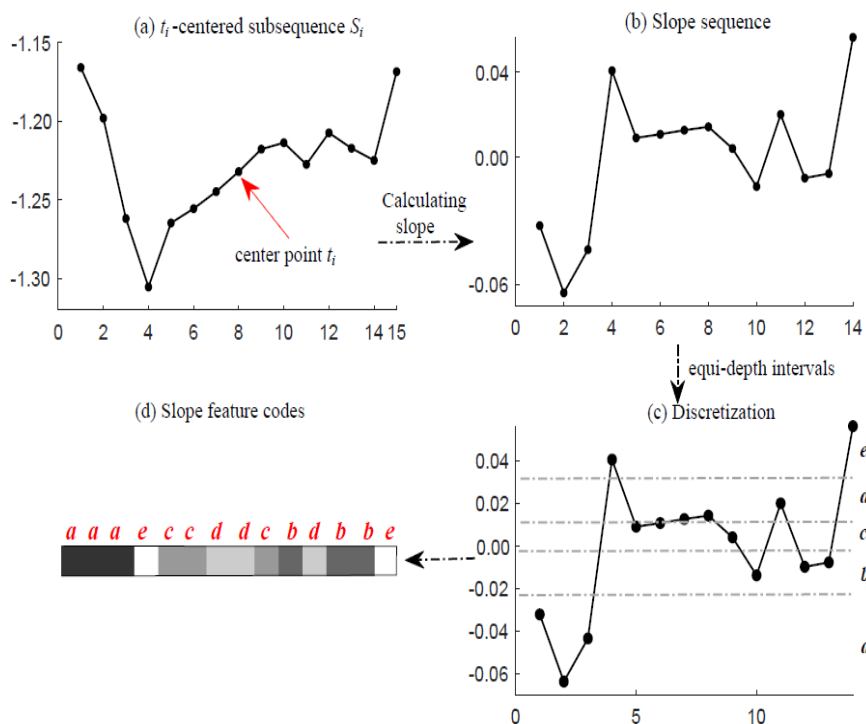
4. 杨皓强（基于注意原型和对抗特征的股票趋势预测，2021届）



5. 张 月 (基于多维时间序列的唇语识别, 2021届)



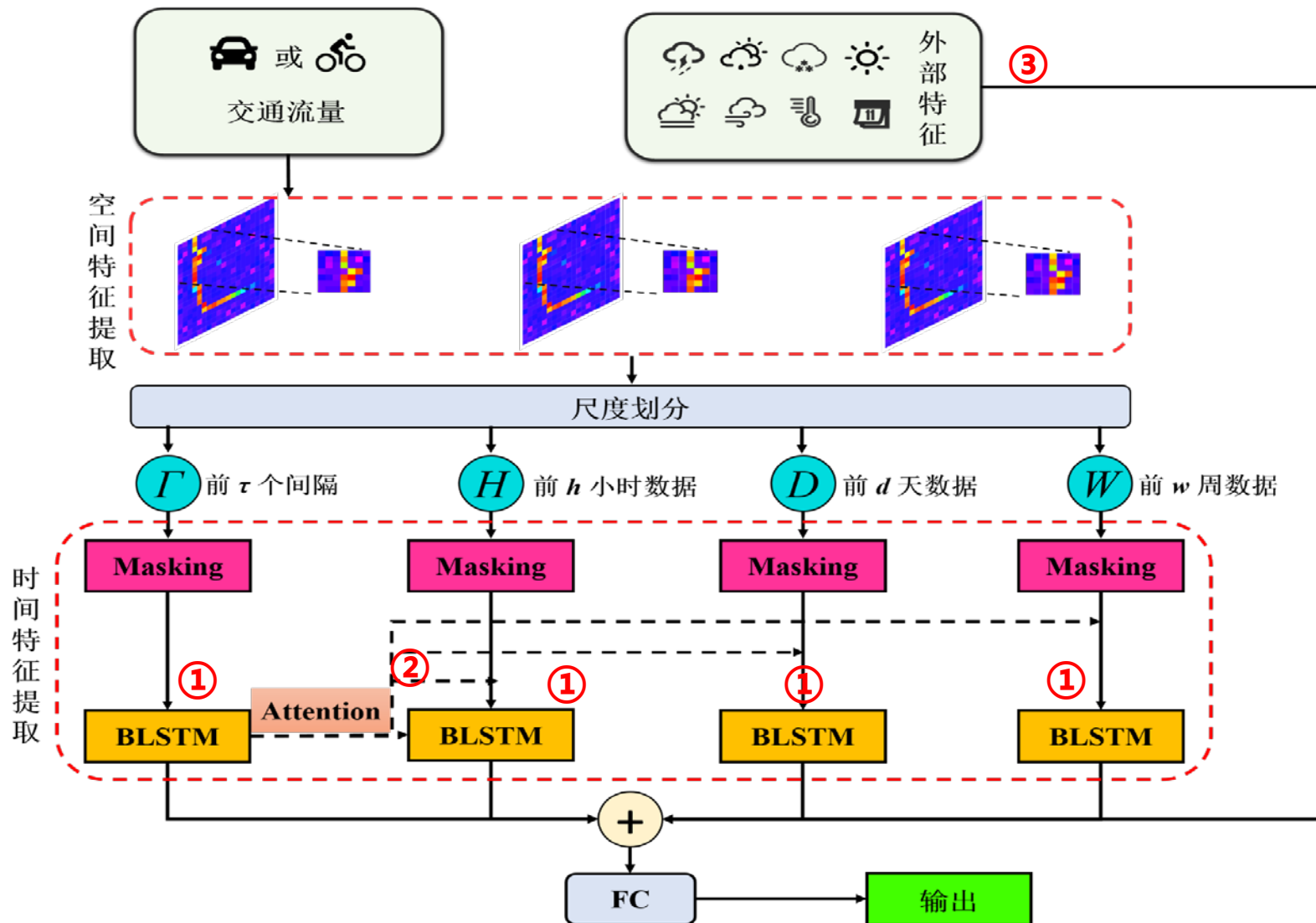
6. 林钱洪（基于形状特征的时间序列分类算法及其应用研究，2020届）



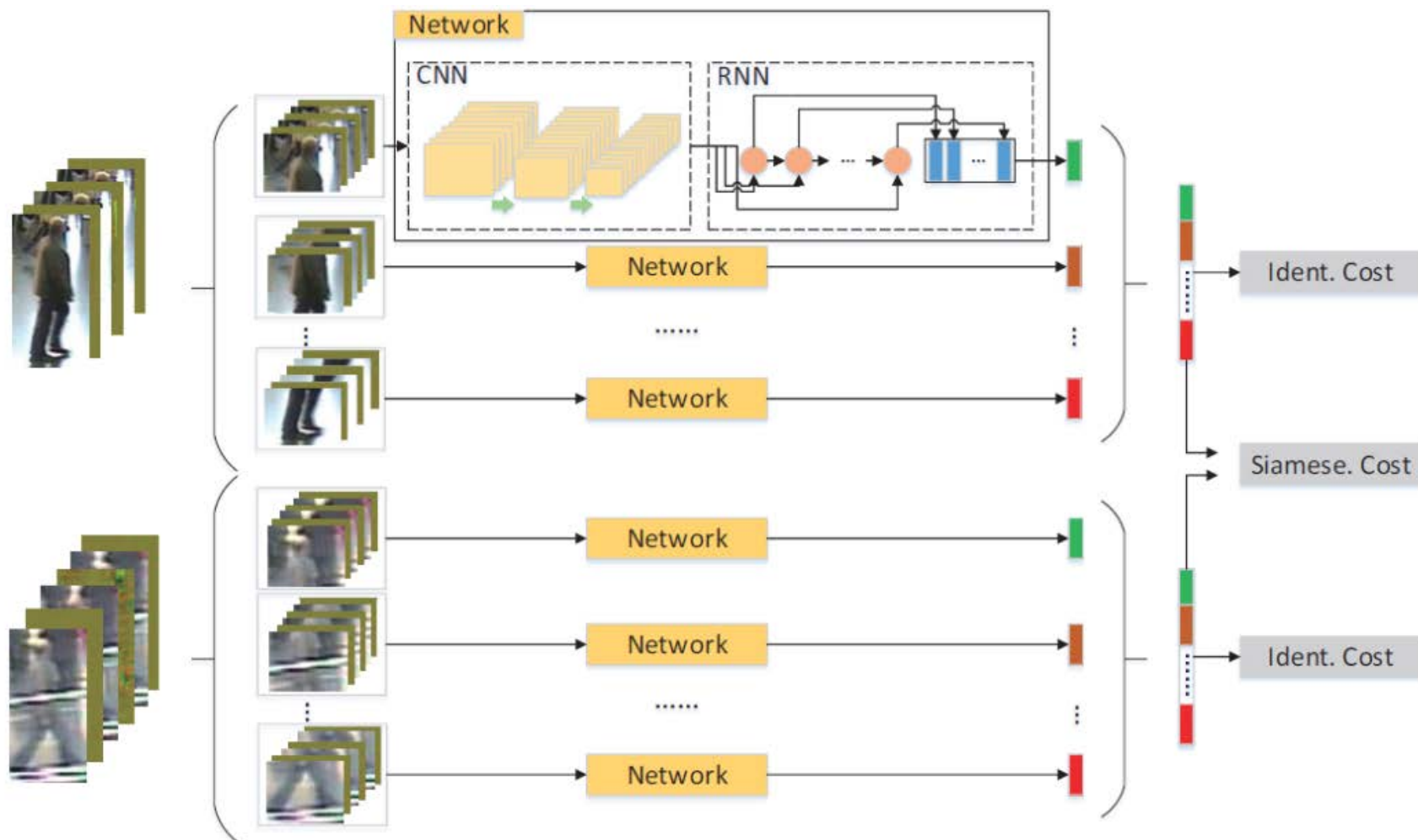
Yuan J, Lin Q, Zhang W, et al. Locally Slope-based Dynamic Time Warping for Time Series Classification[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management(CIKM 2019). 2019: 1713-1722. (CCF-B)

魏池璇, 王志海, 原继东*, 林钱洪. 时间序列可变尺度的时域与频域特征求解及其分类算法[J]. 软件学报, 2022 (CCF A中文)

7. 王冀琛 (基于时空网络的交通流预测, 2020届)



8. 刘 杰 (基于空间时序特征的行人重识别, 2019届)



Jie Liu, Cheng Sun, Xiang Xu, Baomin Xu*, Shuangyuan Yu. A spatial and temporal features mixture model with body parts for video-based person re-identification. Applied Intelligence, 2019, 49: 3436-3446. (SCI-An2期刊)

四、国际交流（国际会议）



五、毕业生的去向

1. 硕士研究生

阿里、百度、快手、字节跳动、京东、银行、移动、联通、
城铁、中铁通号公司等

2. 出国深造

日本大阪大学、澳大利亚悉尼科技大学、美国达特茅斯学院、
美国贝勒医学院、英国兰卡斯特大学、法国约瑟夫傅里叶大学

3. 博士研究生

中国电子科技集团、中国医学科学院、北京交通大学、郑州
大学、桂林电子科技大学、北京建筑大学等科研院所与高校

六、校企合作

inspur 浪潮

Bai du 百度

 中国移动
China Mobile

 蚂蚁金服
ANT FINANCIAL

 华夏银行
HUAXIA BANK

CETC
中国电科

 航天科工

铁科院

七、团队建设



八、联系方式

- 孙永奇：九教西207，13521884358
- 原继东：九教北512，18310028963
- 王志海：九教北502，17090149371
- 徐保民：九教北519，15010510939
- 徐 薇：九教西207，13671103567
- 刘海洋：逸夫楼101，18611249791



欢迎大家线下或进群交流

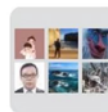
线下交流

- 集中时间：

7月6日下午15:30-18:00

- 其他时间：在群里预约

- 地点：九教西207



群聊：云计算与数据科学研究所
2023 夏令营



该二维码7天内(7月11日前)有效，重新进入将更新

欢迎加入我们团队

掌握云计算、探索大数据！

<http://ccd.bjtu.edu.cn>

