Lecture 1
Introduction

Lin ZHANG, PhD
School of Computer Science and Technology
Tongji University
Fall 2025

Course Info

Contact Information

Room 418L, Jishi Building
TA: Linfei Li, cslinfeili@tongji.edu.cn

All material can be found at

https://cslinzhang.github.io/home/

Note: The first 6 lectures will be given by Lin ZHANG and the remaining 10 lectures will be given by Dr. Wei TANG

Materials

- Major materials
  - My slides
- References
  - 《计算机视觉：原理算法与实践》， 张林等， 2025年5月
  - 《机器学习》， 周志华， 2016
  - 《统计学习方法》（第2版）， 李航， 2019
  - Some papers

Examination

- Homework 30%: 3 times, and each time 10%.
- Paper reading and presentation 20%
  - Read a paper related to machine learning and do a presentation
- Final report and presentation 50%
  - Select a problem related to your research direction, try to solve it with machine learning techniques, write an essay and finally do a presentation
- Being absent >=1/3 lectures, you will fail this course

Arrangement of Lectures (temporarily)

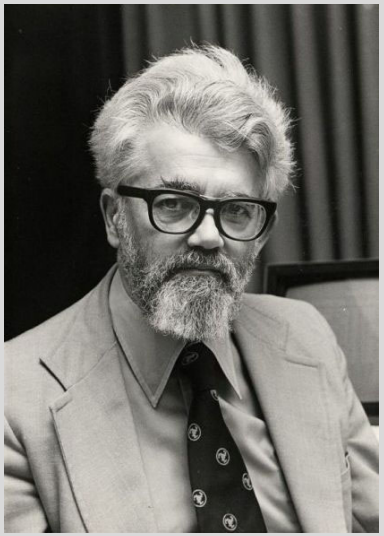
- Basic Concepts and Model Evaluation
- AdaBoost and Cascade Structure
- Principle Component Analysis
- Sparse Representation based Classficiation
- Linear Model
- Neural Network and CNN
- Fundamentals of Convex Optimization
- Support Vector Machines
- Transformer-based Object Detection
- Applications of CNN

A little history about AI

人工智能

1956年，**麦卡锡**召集哈佛大学、麻省理工学院、IBM公司、贝尔实验室的研究人员召开**达特茅斯会议**正式提出“**人工智能**”

2006年达特茅斯会议当事人重聚，左起：**摩尔、麦卡锡、明斯基、赛弗里奇、所罗门诺夫**



John McCarthy

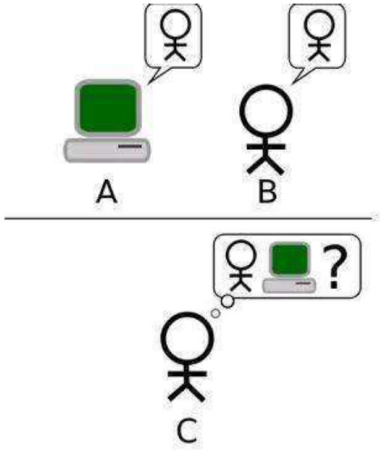
人工智能之父

人工智能是指计算机系统具备的能力，该能力可以履行原本只有依靠人类智慧才能完成的复杂任务

A little history about AI

什么是人工智能?

- 指由人制造出来的机器所表现出来的智能
  - 通常指通过计算机程序来呈现人类智能的技术
- **遗憾的是，“智能”本身难以定义清楚!**
  - 行为定义的智能Behavior defined intelligence
  - 即**图灵测试**定义的智能（不管内涵，只管外延）



什么是人工智能?

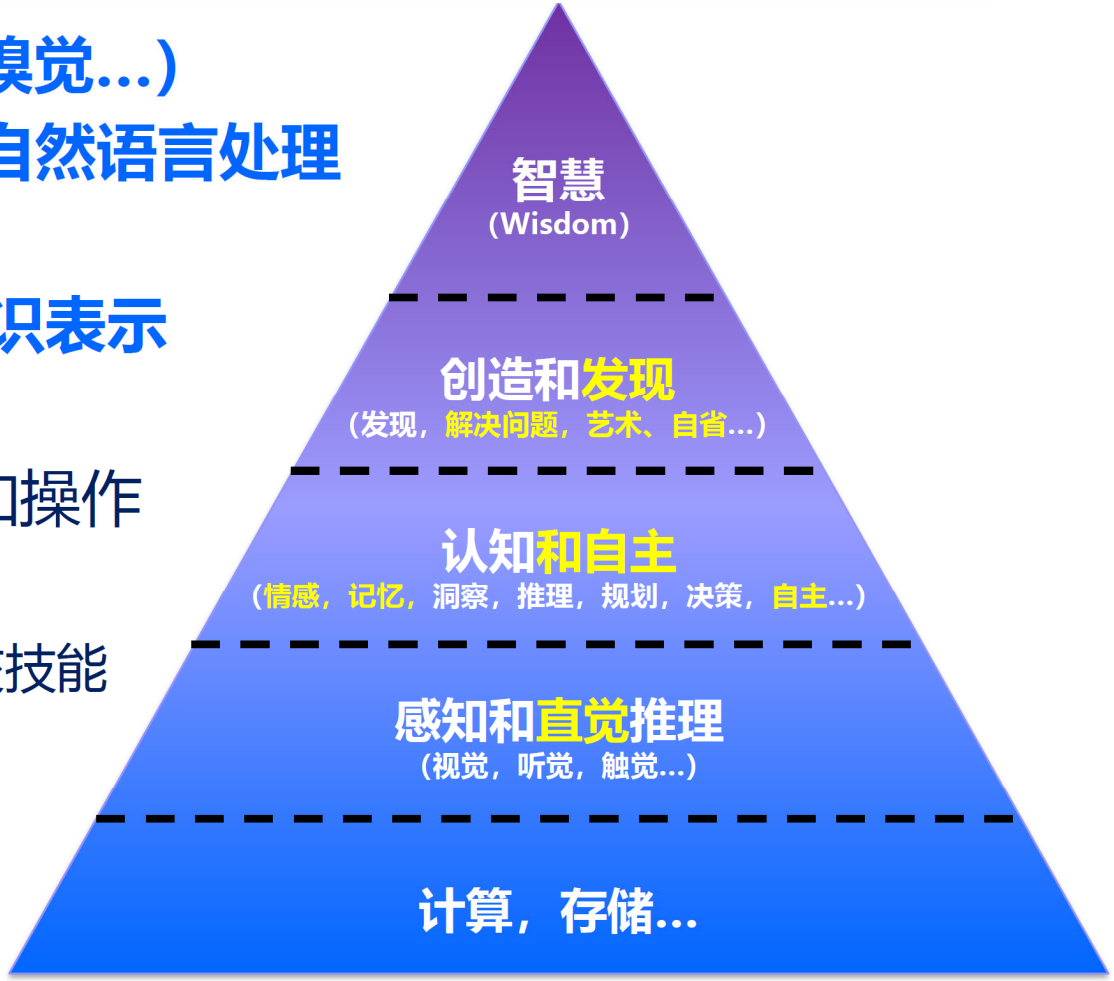
- 行为定义的智能Behavior defined intelligence
  - **系统的表现是智能的**

- 在计算机领域，人工智能是指对“智能代理”的研究
  - 任何可以**感知环境**并**采取行动**以最大可能达成其**特定目标**的任何设备都是智能代理【维基百科】

A little history about AI

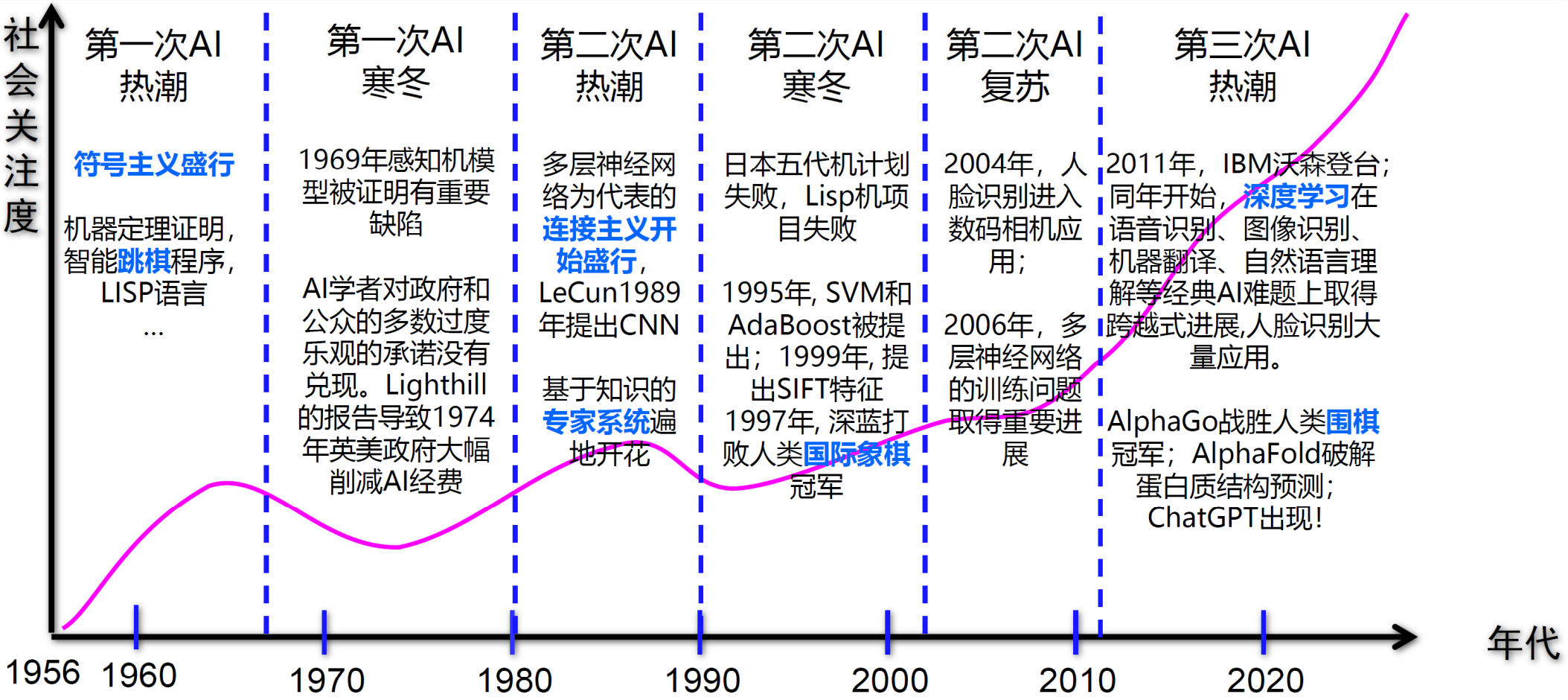
人工智能的内涵和任务

- **Perception感知（视觉，听觉，嗅觉…）**
- **Natural language processing自然语言处理**
- **Learning学习**
- **Knowledge representation 知识表示**
- Planning规划
- Motion and manipulation运动和操作
- Social intelligence社会智能
  - Affective Computing情感计算/社交技能
- Reasoning, problem solving
- Creativity创造力
- General intelligence通用智能



A little history about AI

人工智能发展的历史沿革



符号主义、知识驱动的AI方法论

连接主义、大规模数据驱动的AI方法论

三次人工智能浪潮背后的方法论

- **第一次浪潮：基于符号的推理与搜索**
  - 模拟人的符号推理方式
  - 搜索树：解决迷宫问题
- 符号主义
  - 符号主义的实现基础是纽威尔和西蒙提出的物理符号系统假设
    - 人类**认知和思维的基本单元是符号**，而**认知过程就是在符号表示上的一种运算**。人是一个物理符号系统，计算机也是一个物理符号系统，故可用计算机来模拟人的智能行为，即用计算机的符号操作来模拟人的认知过程
    - 实质就是模拟人的左脑抽象逻辑思维，通过**研究人类认知系统的功能机理，用符号之间的逻辑关系来描述人类的认知过程**，并把这种符号输入到能处理符号的计算机中，就可以模拟人类的认知过程，从而实现人工智能

三次人工智能浪潮背后的方法论

- **第一次浪潮：基于符号的推理与搜索**
  - 模拟人的符号推理方式
  - 搜索树：解决迷宫问题
- 符号主义
  - 太乐观了【 承诺太多，最终做不到 】！
    - 解决不了更复杂的现实问题
      搜索空间太大了：如围棋
    - **大量问题难以转化为符号推理问题**
      例如：人脸识别、语音识别等模式识别问题（非结构化数据的结构化）

三次人工智能浪潮背后的方法论

- **第二次浪潮：依赖人类符号化知识的专家系统**
  - **以依赖符号化知识库和符号推理的专家系统为主**
  - 知识表示是人工智能的核心难题
  - 人工智能研究早期主流的知识表示方法——**符号主义的知识观**
    - 从符号主义的观点来看，**认知就是符号的处理过程**，是智能的基础
    - **符号化的知识表示、推理、运用**是人工智能的核心
    - 知识表示：采用符号表示（实体、关系等）所有知识
    - 知识推理：推理是采用启发式知识及启发式搜索对**问题求解的过程**，其过程可以用某种**形式化的语言**来描述，因而有可能建立起基于符号化知识的人类智能和机器智能的**同一理论体系**

三次人工智能浪潮背后的方法论

- **第二次浪潮：依赖人类符号化知识的专家系统**
  - 多数AI系统建立在符号基础上的知识表示（知识库、知识图谱）
    - 例：医疗辅诊系统，症状->疾病
  - 巅峰之作：IBM 的沃森自动问答系统
    - 2011年，IBM 沃森在问答竞赛 《 危险边缘 》 （Jeopardy）上击败人类
    - 类似问题：地球上最北端的机场是哪个?
    - 背后的技术

      自然语言处理、消息检索、知识表示、自动推理、机器学习等开放式问答技术

三次人工智能浪潮背后的方法论

- **第二次浪潮：依赖人类符号化知识的专家系统**
  - 还是太乐观了
    - 雄心勃勃的日本**第五代计算机**计划失败
    - 美国Cyc常识知识库项目陷入困境（至少不算成功）

      1984年启动，Douglas Lenat教授领衔，以手工建立知识库为主，包含了320万条人类定义的断言，涉及30万个概念，15000个谓词
    - **挑战性问题：常识是否可穷尽枚举?**
    - 难以解决复杂的现实问题

      知识表示困境：文字识别尚可，人脸识别用什么知识表示?
  - 方法论层面
    - 方法论上的悄然变迁，基于**专家知识人工设计特征**，采用**统计模式识别**和**机器学习**（包括神经网络）工具，学习较小规模数据之间统计关系，成为主流方法

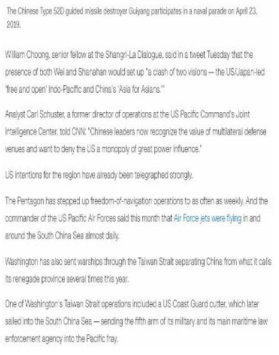
三次人工智能浪潮背后的方法论

- **第二次浪潮：依赖人类符号化知识的专家系统**
  - 第二次浪潮末期：数据驱动的机器学习方法的崛起
  - 基本原理—基于函数拟合的预测问题
    - 用[较大量的]成对的(x_i, y_i)数据，拟合一个带有θ参数的函数f
      - 本质：学习x和y的相关性；类比：学生学习过程，x_i是**考题**，y_i是**答案**
    - 函数f经常是人工设计的，例如：线性函数$y = f(x) = Ax$
    - **参数θ**量相对较少（但也经常在数十万甚至数百万量级）

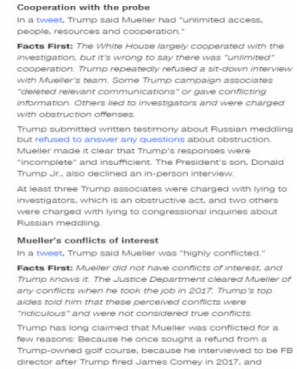
A little history about AI

三次人工智能浪潮背后的方法论

- **第二次浪潮：依赖人类符号化知识的专家系统**
  - 方法论层面——**基于知识的特征设计**
    - **两篇文章**的相似度计算

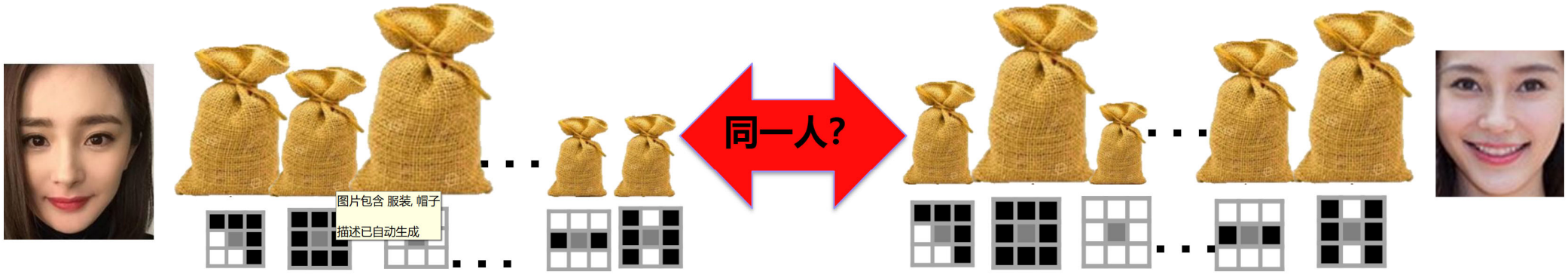


同一篇文章?



三次人工智能浪潮背后的方法论

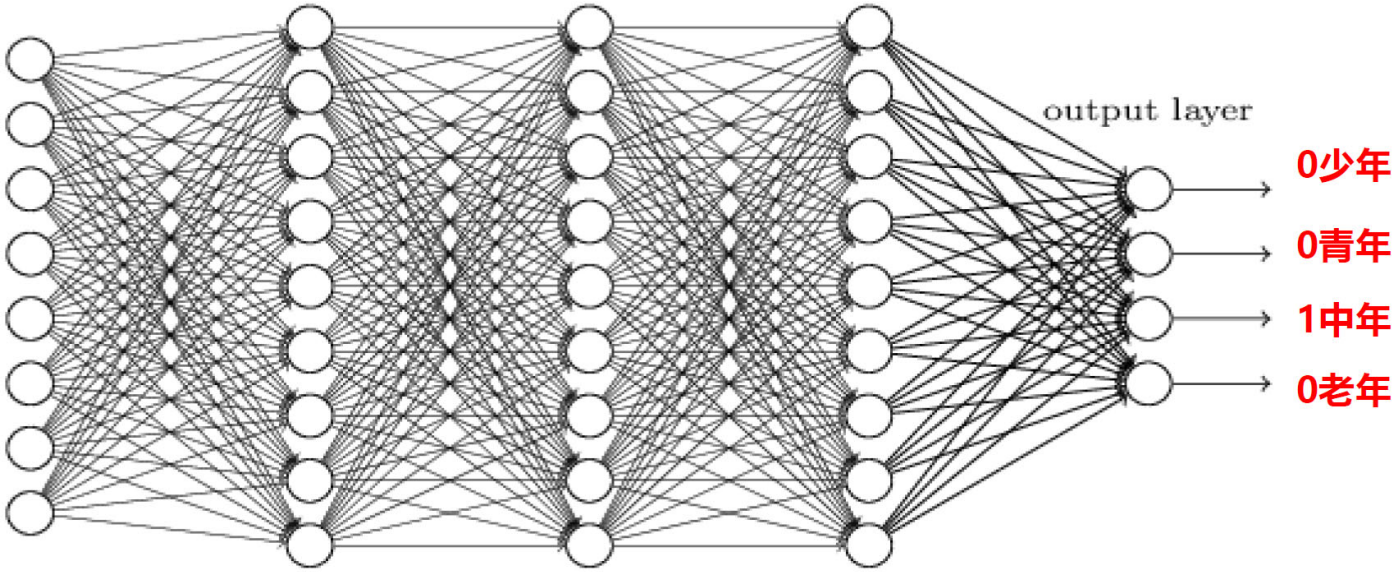
- **第二次浪潮：依赖人类符号化知识的专家系统**
  - 方法论层面——**基于知识的特征设计**
    - **两张人脸**的相似度计算
      - **步骤1**：图像中若干个点形成的**微模式类型**
      - **步骤2**：统计人脸上**不同微模式的出现频次**作为不同人脸的**特征表示**



三次人工智能浪潮背后的方法论

- **第三次浪潮：依赖大量数据的深度学习方法**
  - 用神经网络作为映射函数f**直接学习**从输入x预测输出y
    - **较少依赖人工设计**
  - **题海战术（动辄百万，千万量级）**

x

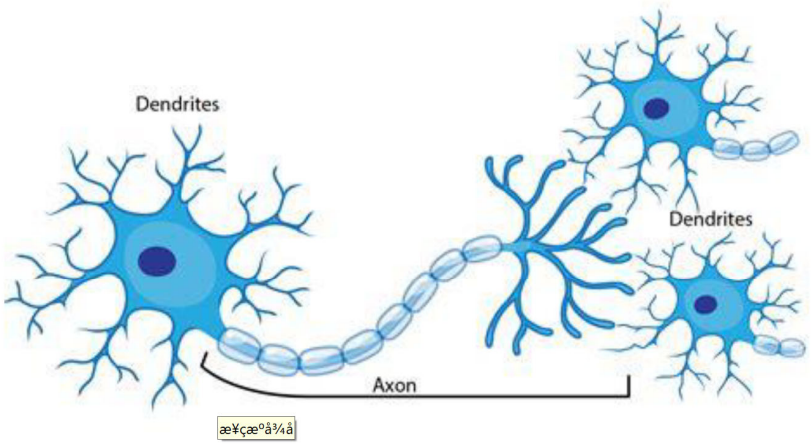


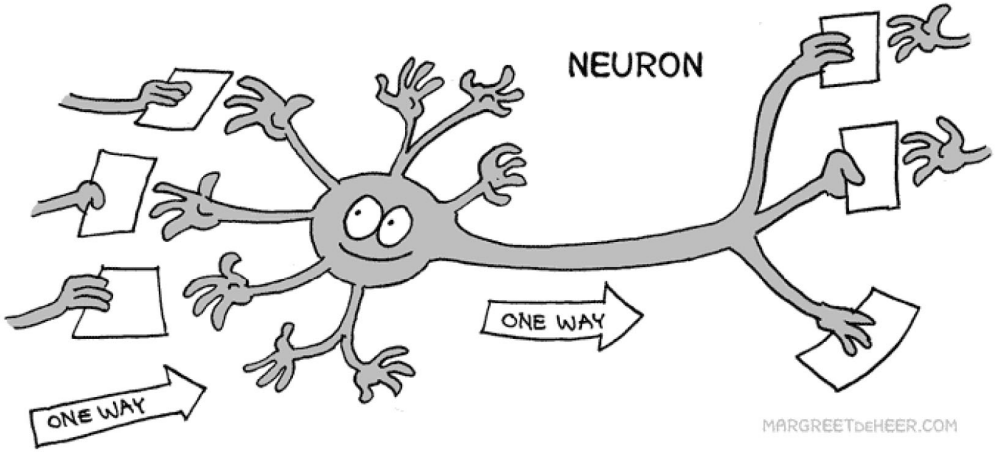
三次人工智能浪潮背后的方法论

- **第三次浪潮：依赖大量数据的深度学习方法**

深度学习(深度卷积神经网络)的缘起

- 生物脑中的神经网络，单个神经元的**功能**
  - 接收前面神经元的输入，汇总➔决策➔传递





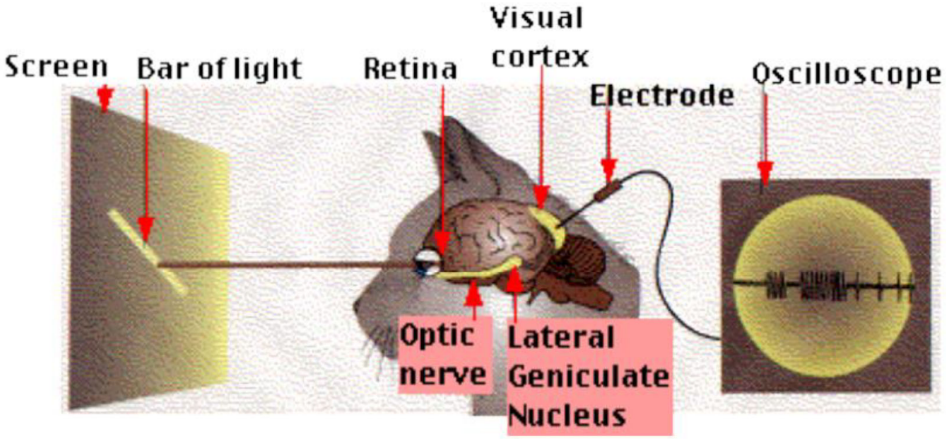
A little history about AI

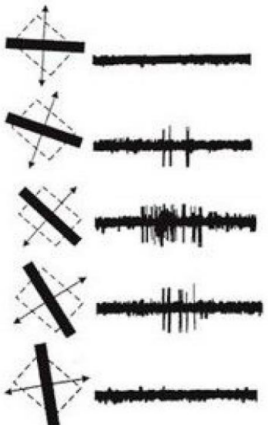
三次人工智能浪潮背后的方法论

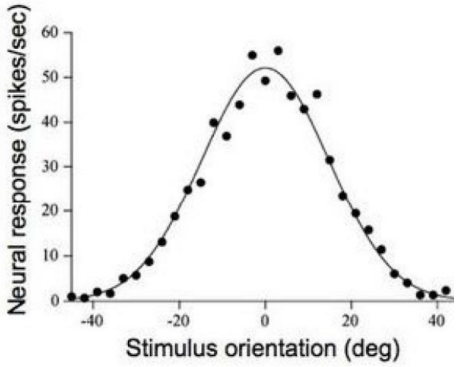
- **第三次浪潮：依赖大量数据的深度学习方法**

深度学习(深度卷积神经网络)的缘起

- 生物脑中的神经网络，单个神经元的**功能**
  - 接收前面神经元的输入，汇总➔决策➔传递
  - 初级视觉皮层(V1)区简单细胞
    - **功能是检测不同朝向的线段** Hubel & Wiesel, 1959, 1962, …





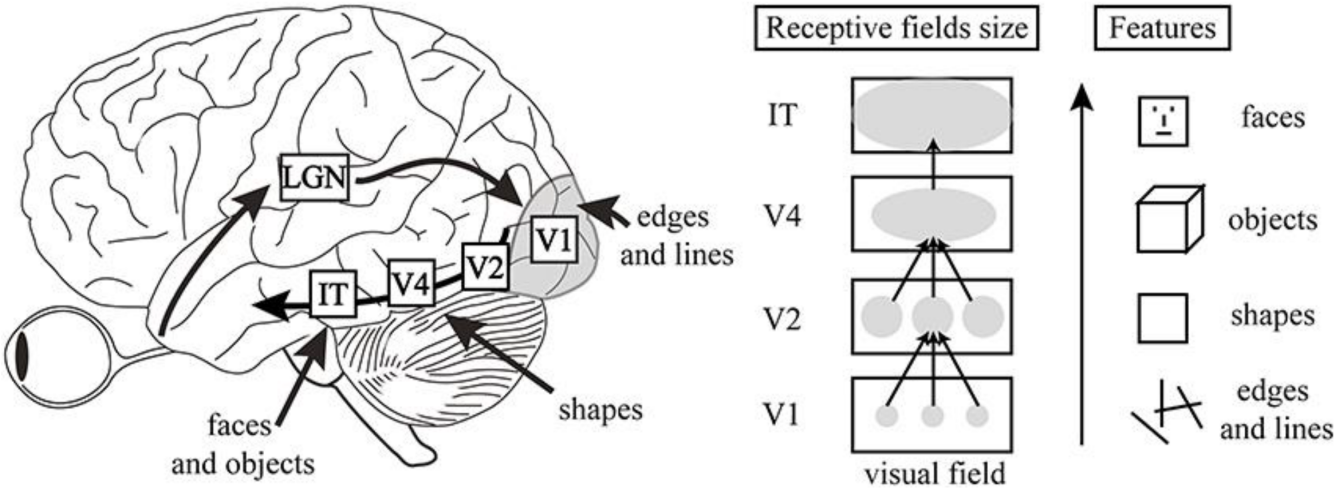


三次人工智能浪潮背后的方法论

- **第三次浪潮：依赖大量数据的深度学习方法**

深度学习(深度卷积神经网络)的缘起

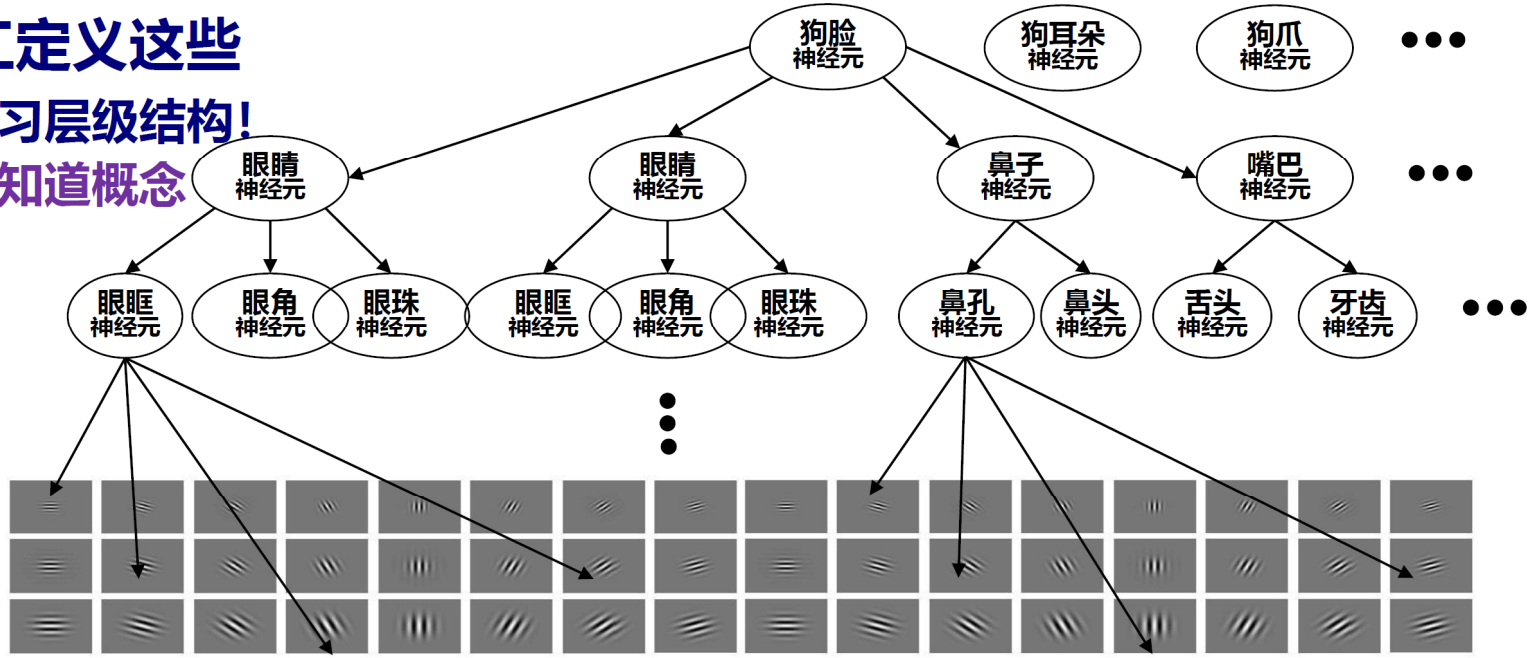
- 生物脑中的神经网络，**大量神经元互联**
  - 视觉通路神经细胞层级感受野假设
    - 响应越来越复杂的模式➔祖母细胞理论
    - 可见越来越大的(视网膜)感受野：**类比从普通士兵到总司令**



三次人工智能浪潮背后的方法论

- **第三次浪潮：依赖大量数据的深度学习方法**

  - **所谓深度学习主要是指多层神经网络**
    - **一个例子：多层神经网络怎么找到狗?**
    - **但不需要人工定义这些**
      - **算法自动学习层级结构!**
      - **算法不需要知道概念**

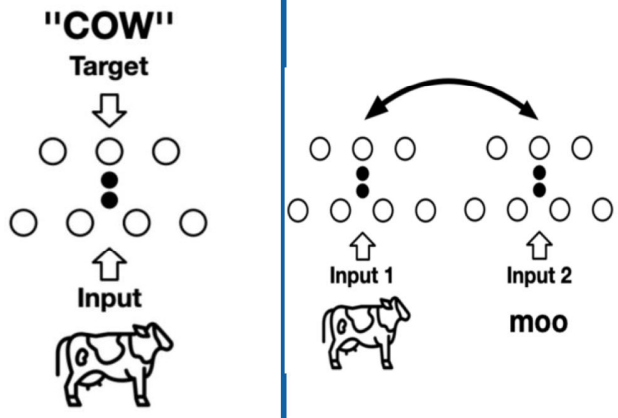


A little history about AI

预训练大模型有望突破通用人工智能瓶颈

开启了基于自监督学习的“大数据+大模型”新范式，从大规模的无标注数据中挖掘隐含的监督信息进行通用知识学习，成为迈向通用人工智能的重要途径

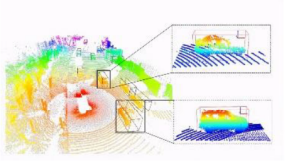
1 从有监督到自监督



有限人工标注训练样本

大规模无标注训练样本

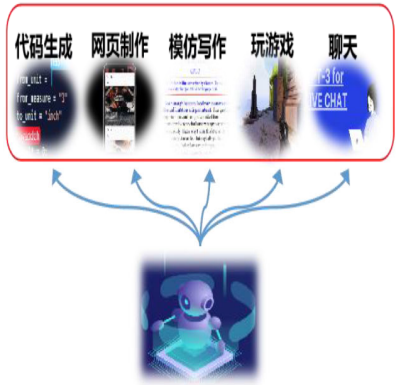
2 从专用小模型到通用大模型



目标检测

人脸识别

高度定制



大模型

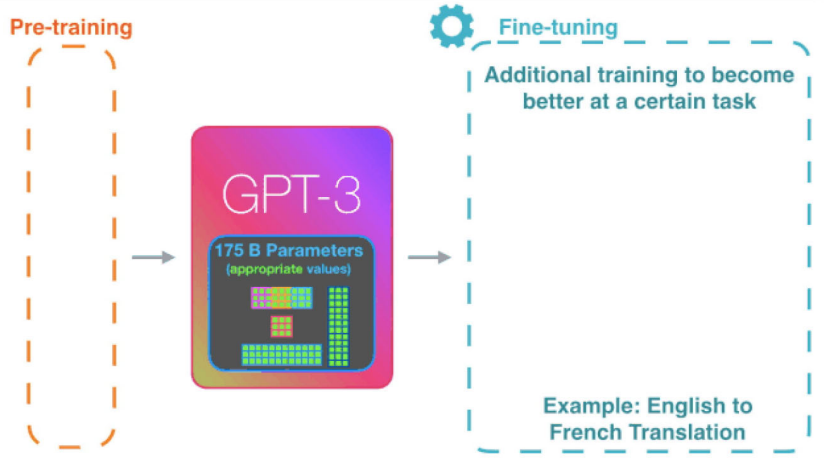
多专多能

A little history about AI

预训练大模型有望突破通用人工智能瓶颈

自然语言理解领域的大杀器——GPT-3

- **2020年6月，OpenAI GPT-3**
  - 1750亿参数，比其前身多100倍
    - 比之前最大NLP模型要多10倍
    - 花费460万美元进行训练
  - 大力出奇迹：见过巨量的人类语言
    - 训练语料：**3000亿单词（tokens）**
      - 60%：C4语料库（爬虫项目 Common Crawl 在 2019 年 4 月全网部分文本快照）
      - 22%：WebText2 （OpenAI自己收集的，未全部开放）
      - 16%：Books
      - **3%：Wikipedia**
    - 整个英语维基百科（约600万个词条）仅占其训练数据的3%

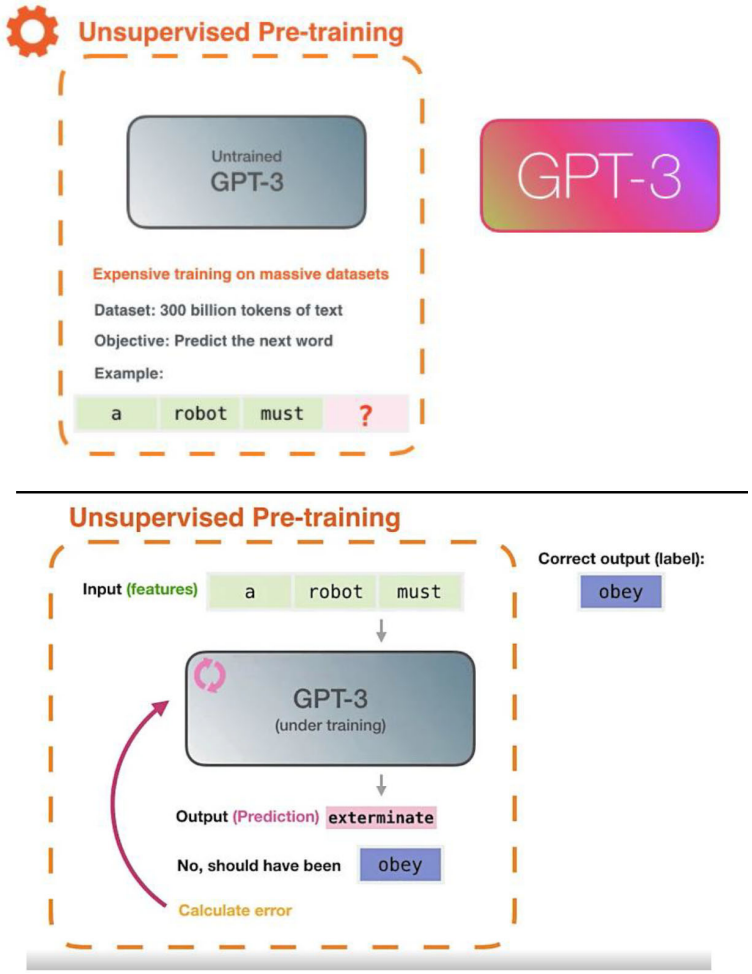


A little history about AI

预训练大模型有望突破通用人工智能瓶颈

自然语言理解领域的大杀器—GPT-3预训练

- GPT-3 预训练
  - 无监督学习（自监督学习）
  - Language Modeling：从前述词预测下一个词
    - 如左图中通过 “a robot must” 来预测下一个词“obey”
  - 通过学习语言，同时学到了以自然语言中表达的大量“知识”



来源：How GPT3 Works - Visualizations and Animations, Jay Alammar

预训练大模型有望突破通用人工智能瓶颈

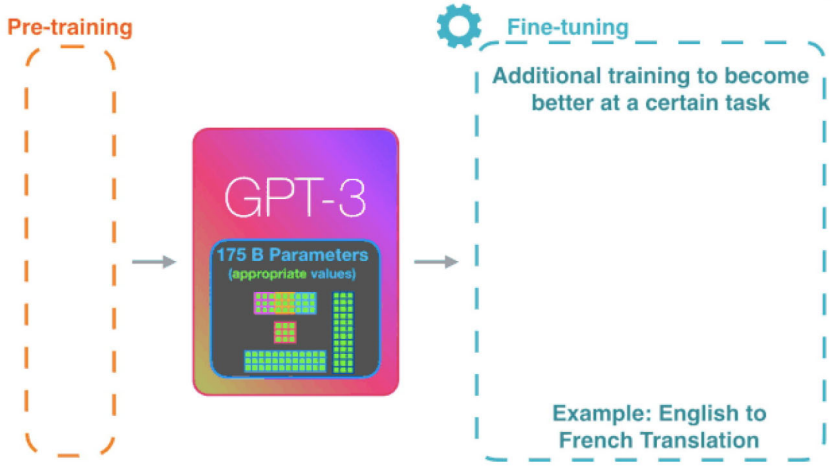
GPT, GPT-2, GPT-3

	GPT	GPT-2	GPT-3
数据集	5GB: BookCorpus	40GB: WebText	45TB: Common Crawl, WebText2, Books1, Books2, Wikipedia
参数量	117M	1.5B	175B
训练方法	Unsupervised pre-training, fine-tuning on each task	Unsupervised multitask pre-training via meta-learning, zero shot	Unsupervised multitask pre-training via meta-learning, zero/one/few shot
模型结构	Decoder (layer=12, dim=768, head=12)	Decoder (layer=48, dim=1600)	Decoder (layer=96, dim=12888, head=96)

预训练大模型有望突破通用人工智能瓶颈

自然语言理解领域的大杀器——GPT-3

- **2020年6月，OpenAI GPT-3**
  - 1750亿参数，比其前身多100倍
    - 比之前最大NLP模型要多10倍
    - 花费460万美元进行训练
  - 大力出奇迹：见过巨量的人类语言
    - 整个英语维基百科（约600万个词条）仅占其训练数据的3%



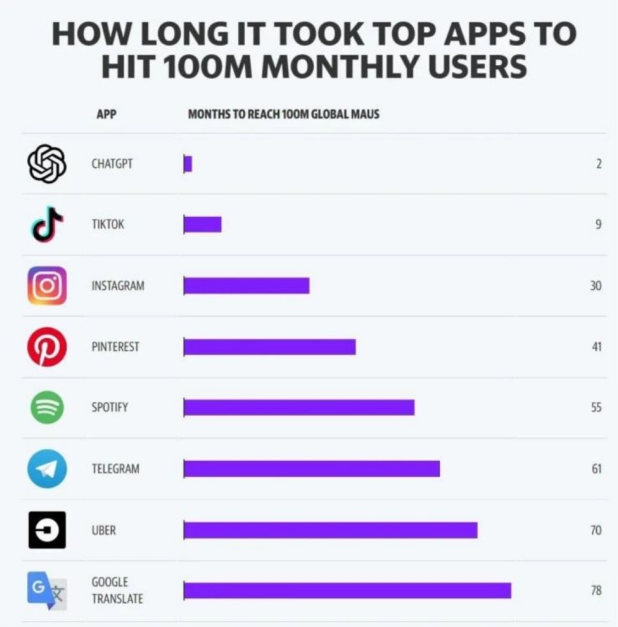
- 可以做什么？
  - 回答问题，基于问题的搜索引擎，聊天机器人，机器翻译，续写文章...

A little history about AI

预训练大模型有望突破通用人工智能瓶颈

ChatGPT是什么?

- ChatGPT基于大规模语言模型GPT3.5，通过**人类反馈学习微调**而来的对话生成大模型。不再是传统意义的人机对话系统，是以自然语言为交互的通用语言处理平台
- **超出预期的交互体验**
  - 通用的意图理解能力
  - 强大的连续对话能力
  - 智能的交互修正能力
  - 较强的逻辑推理能力



推出**2个月**即达到**1亿**活跃用户
历史上**增长最快**的消费者应用程序

将对文字编辑、程序编译、智能问答等
行业带来巨大冲击

A little history about AI

预训练大模型有望突破通用人工智能瓶颈

ChatGPT基础数据：文本与代码

该页slide来自中科院自动化所刘静研究员

2020年 OpenAI利用 **45T** 文本数据，通过自监督训练获得基础大模型**GPT-3**，实现**流畅性、知识性**

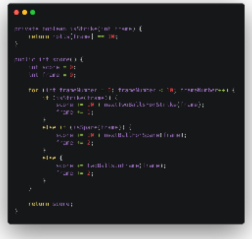
专业书籍
维基百科
互联网文本
……

丰富的语言知识
多样的语言表达

GPT-3
能说会道

更多更新数据

C++
Java
Python
……



全面的逻辑实现
详细的代码注释

CodeX
逻辑编程

更多更新代码

2022年 OpenAI利用更多更新文本数据和代码数据的混合学习，得到更强的基础大模型**GPT-3.5**，成为ChatGPT的基础模型，实现了**流畅性、知识性和逻辑性（推理能力）**

2021年 OpenAI在GPT-3基础上利用 **179G** 代码数据，通过自监督训练获得**逻辑**编程模型**Codex**

A little history about AI

预训练大模型有望突破通用人工智能瓶颈

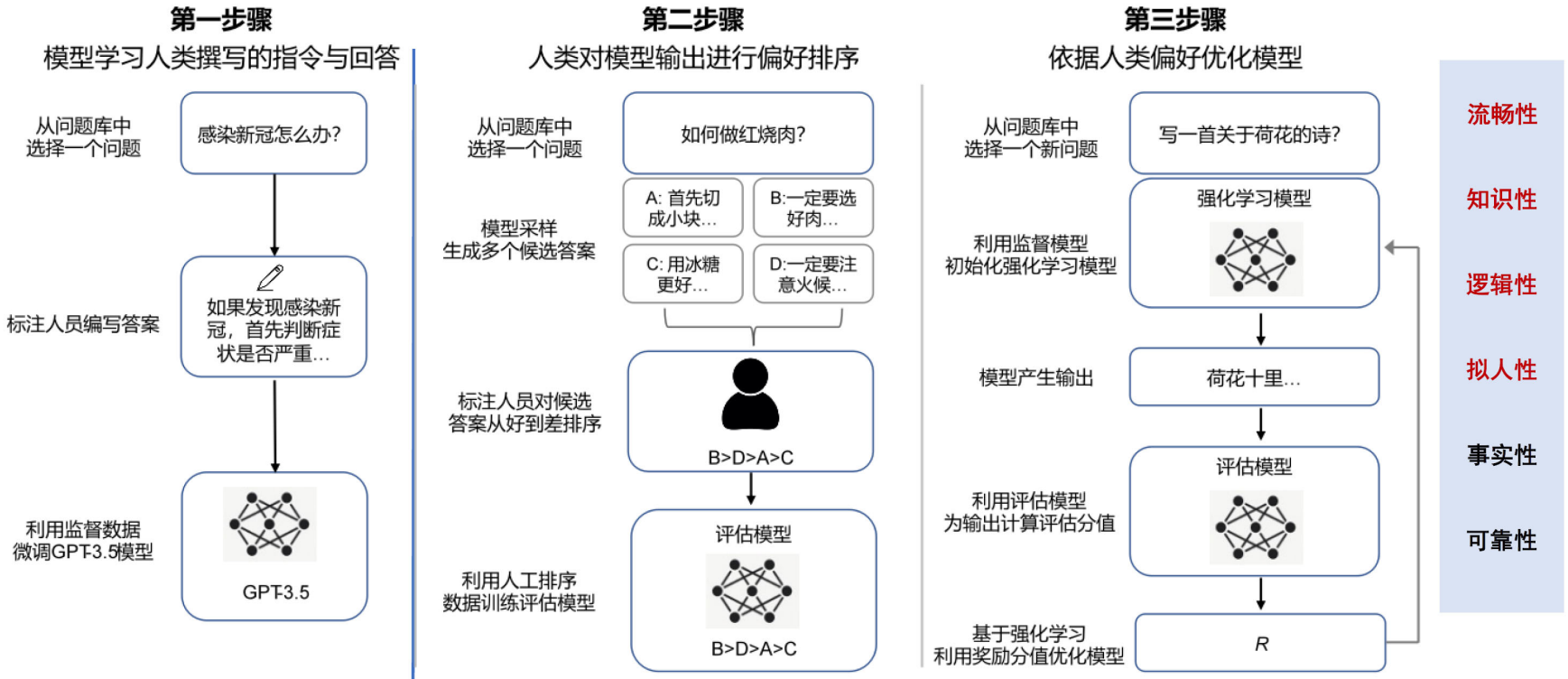
ChatGPT的工作原理

该页slide的来自中科院自动化所刘静研究员

- ChatGPT是通过对话交互方式，对语言大模型文本理解与生成能力的集成展示

OpenAI

GPT 3.5/4
文本大模型



流畅性
知识性
逻辑性
拟人性
事实性
可靠性

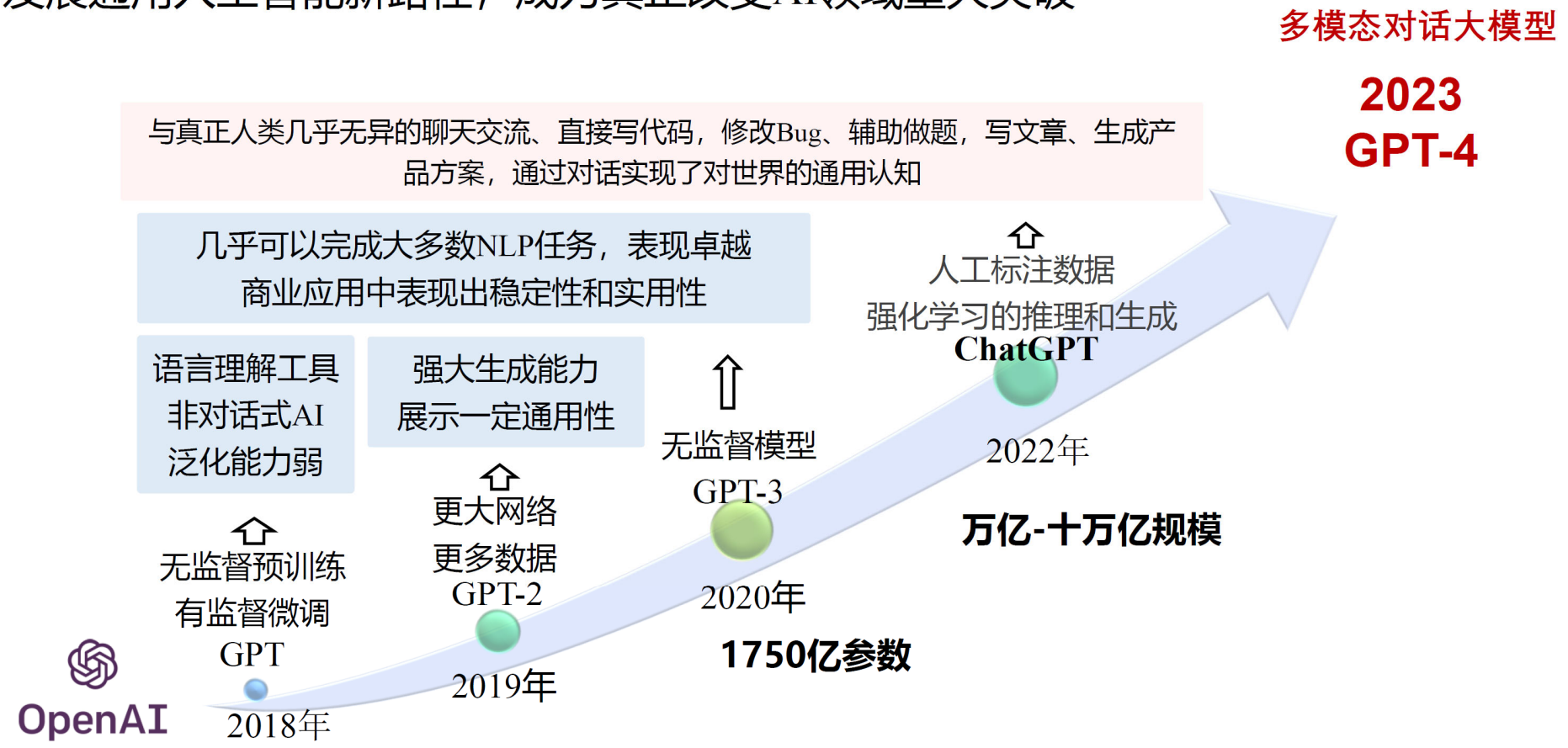
自监督学习	有监督学习 - 指令微调	有监督学习 - 强化学习
强大的文本理解与生成能力	“用户在问什么？”	“用户想要的答案是什么？”

A little history about AI

预训练大模型有望突破通用人工智能瓶颈

ChatGPT以产品为导向，众多技术与成果的集大成者

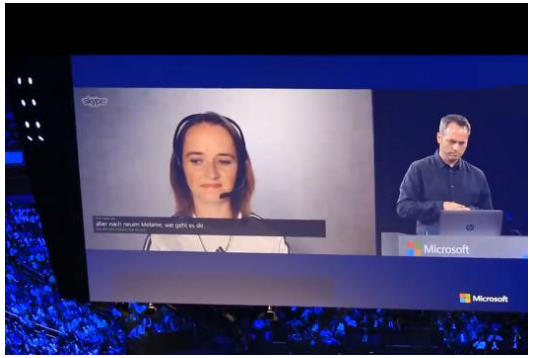
- **大模型技术**与**人类反馈强化学习**融合，实现知识逻辑涌现和人类价值观模拟，探索出了发展通用人工智能新路径，成为真正改变AI领域重大突破



A little history about AI

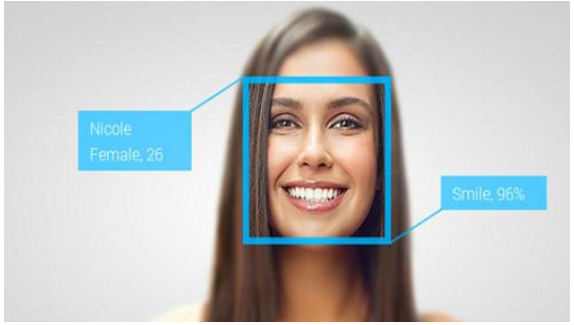
人工智能产业发展加速明显

自然语言处理（NLP）：
微软Skype Translator实现同声传译



计算机视觉（CV）：
格林深瞳的视频监控可智能识别犯罪

计算机视觉（CV）：
Face++的人脸识别云服务



感知、规划和决策：
Google无人驾驶汽车

A little history about AI

人工智能成为世界焦点

人工智能目前已经成为世界各国关注的焦点。2017年7月，中国政府发布了“新一代人工智能发展规划”

人工智能是开启未来智能世界的秘钥，是未来科技发展的战略制高点；谁掌握人工智能，谁就将成为未来核心技术的掌控者

Our work related to ML

Palmprint and palmvein recognition

Lin Zhang, Lida Li, Anqi Yang et al., "Towards contactless palmprint Recognition: A novel device, a new benchmark, and a collaborative representation based identification approach", Pattern Recognition, vol. 69, pp. 199-212, 2017

Our work related to ML

Palmprint verification on mobilephones

Yingyi Zhang, Lin Zhang* et al., Pay by showing your palm: A study of palmprint verification on mobile platforms, in: Proc. ICME, pp. 862-867, 2019.

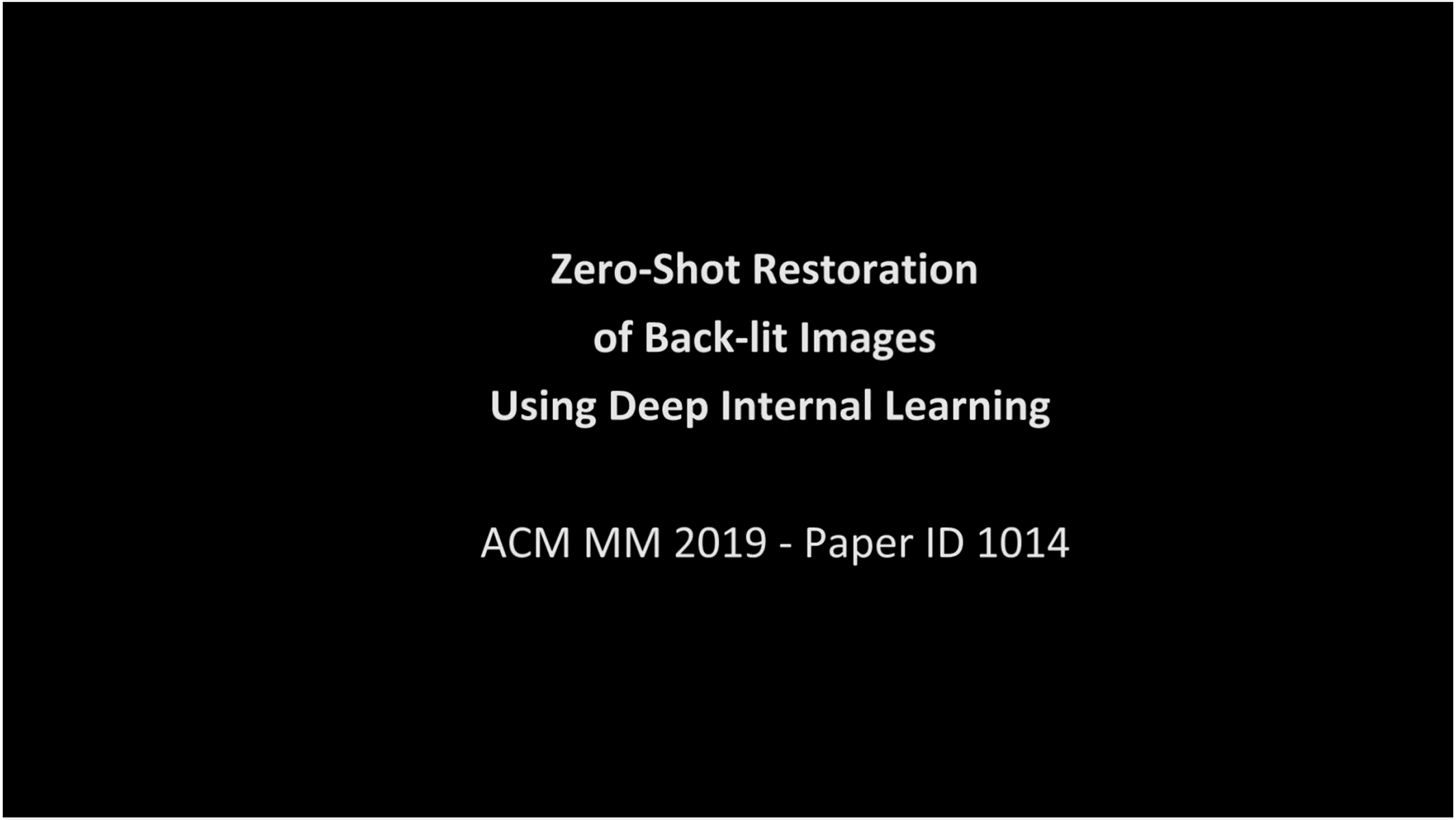
Our work related to ML

图像曝光度的自动校正

Our work related to ML

图像曝光度的自动校正

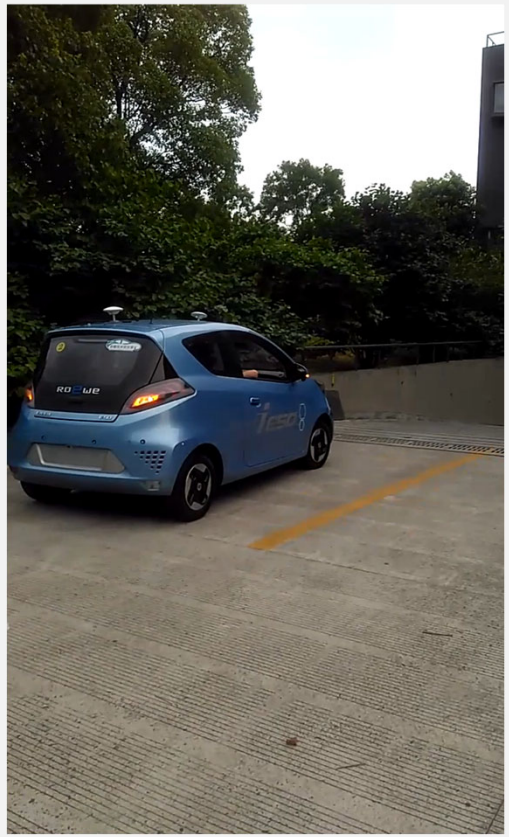
Our work related to ML



Lin Zhang, Lijun Zhang et al., "Zero-Shot Restoration of Back-lit Images Using Deep Internal Learning", ACM Int'l Conf. Multimedia, 2019

Our work related to ML

Short-range Self-parking



Lin Zhang, Junhao Huang et al., "Vision-based parking-slot detection: A DCNN-based approach and a large-scale benchmark dataset", IEEE Trans. Image Processing, vol. 27, no. 11, pp. 5350-5364, 2018.

Our work related to ML

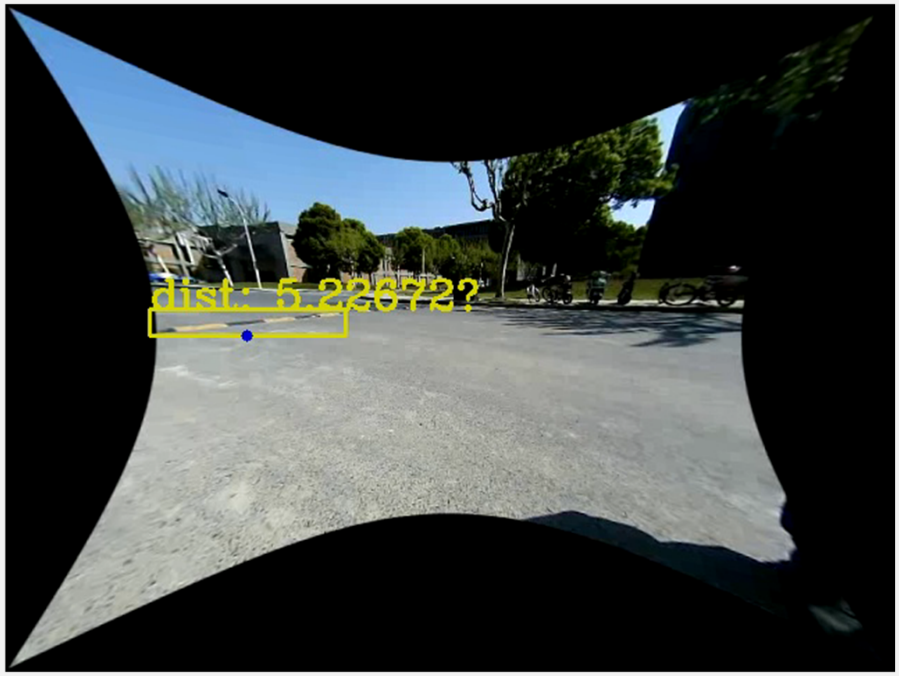
2018年9月，人民网采访报道

在学术界和工业界都产生了较大影响

- 在学术界的影响
  - ✓ CSDN、知乎等多家网络技术媒体对申报人的泊车位检测技术进行了大篇幅介绍和正面评价
  - ✓ 公开的数据集和工具已经被美国南德州大学、韩国汉阳大学、北交大、华南理工、湖南大学等多所研究机构的人员下载使用
- 在工业界的影响
  - ✓ 华为、科大讯飞、纵目、天瞳威视等企业下载使用了我们的数据或复现了我们的算法

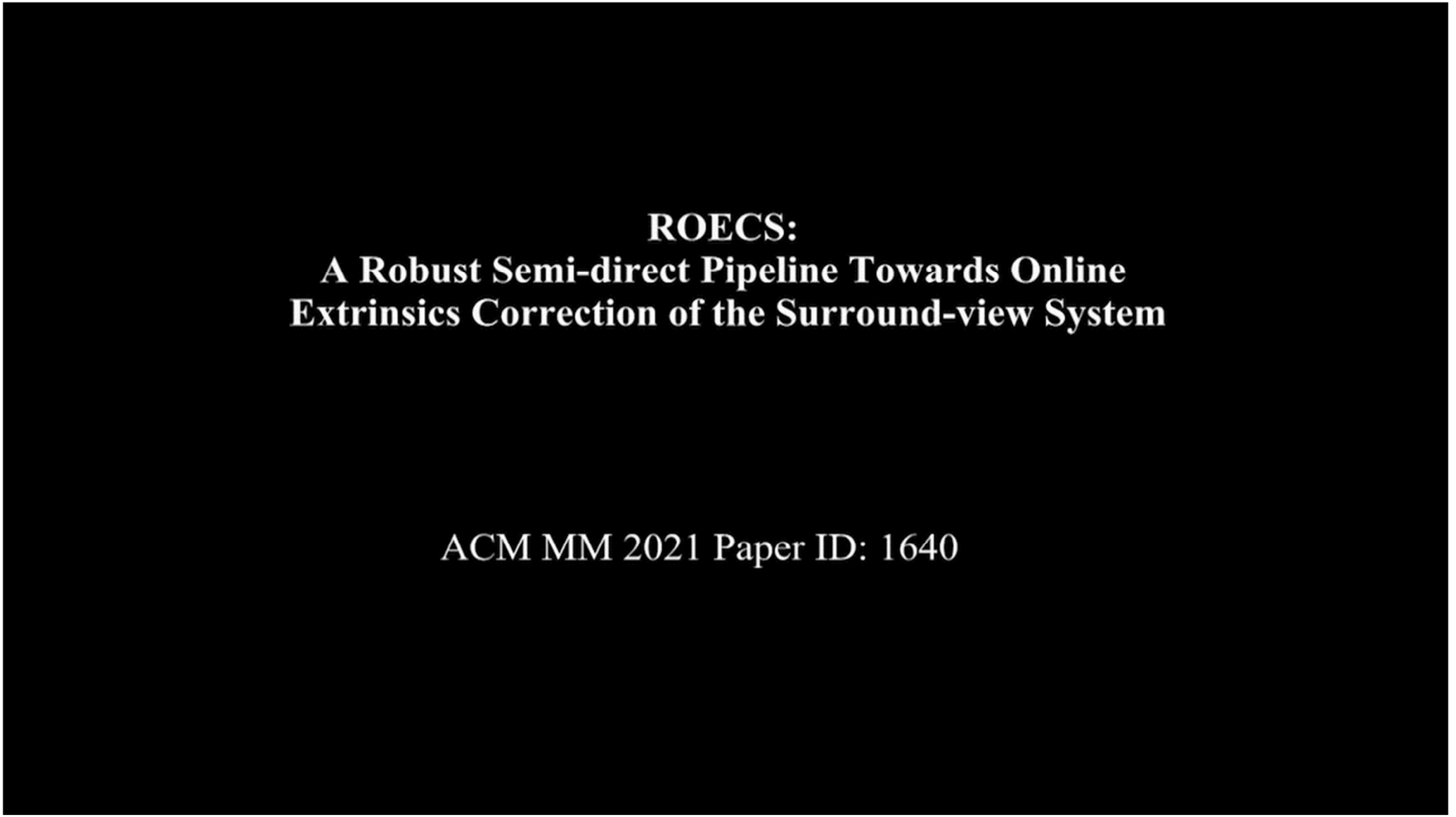
Our work related to ML

Pedestrian and speed-bump detection and distance measurement



Our work related to ML

Online optimization of camera poses in a surround-view system



Tianjun Zhang, Lin Zhang* et al., "ROECS: A Robust Semi-direct Pipeline Towards Online Extrinsics Correction of the Surround-view System", in Proc. ACM Int'l Conf. Multimedia, 2021.

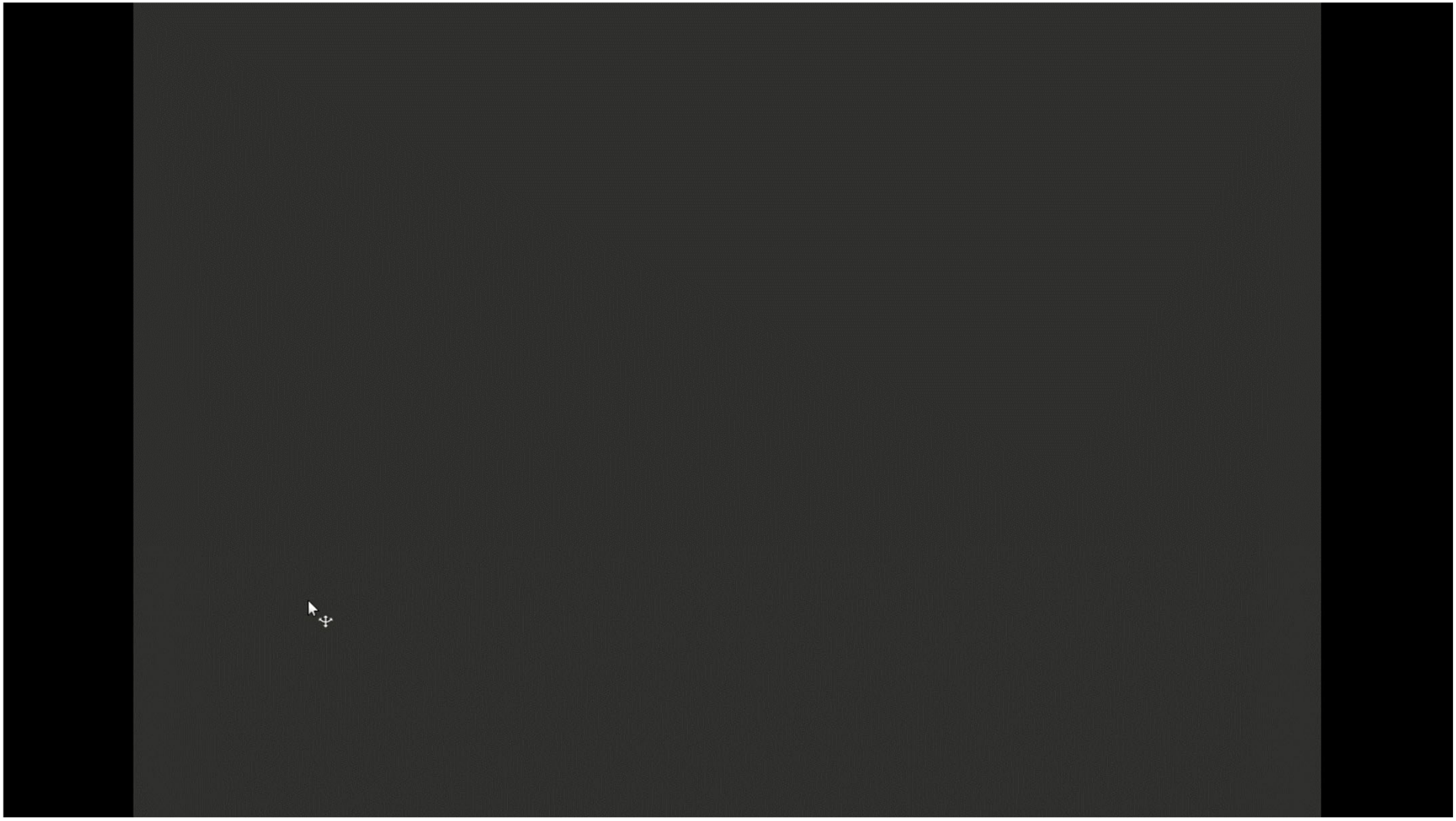
Our work related to ML

SLAM for indoor parking environments

Xuan Shao, Lin Zhang* et al., "MOFIS$_{SLAM}$: A multi-object semantic SLAM system with front-view, inertial and surround-view sensors for indoor parking", IEEE Trans. CSVT, 2022

Our work related to ML

Multi-agent mapping



Tianjun Zhang, Lin Zhang* et al., "CVIDS: A collaborative localization and dense mapping framework for multi-agent based visual-inertial SLAM," IEEE Transactions on Image Processing, vol. 31, 2022

Our work related to ML

Sound localization in noisy environments

Zhanbo Shi, Lin Zhang* et al., Audio–Visual Sound Source Localization and Tracking Based on Mobile Robot for The Cocktail Party Problem, Appl. Sci., 13(10), 2023

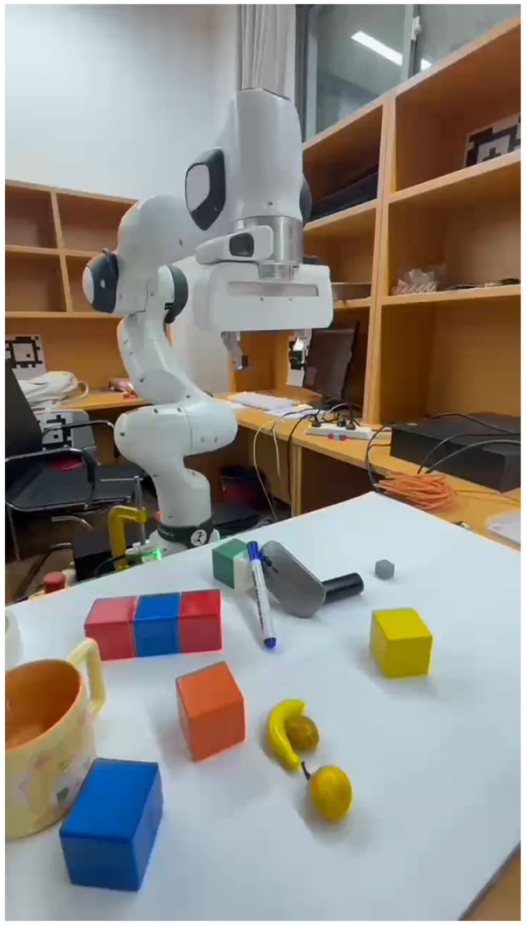
Our work related to ML

Human pose tracking using sparse IMUs

Kaixin Chen, Lin Zhang* et al., "Skeleton-aware Graph-based Adversarial Networks for Human Pose Estimation from Sparse IMUs," ACM Transactions on Multimedia Computing, Communications, and Applications, 2024.

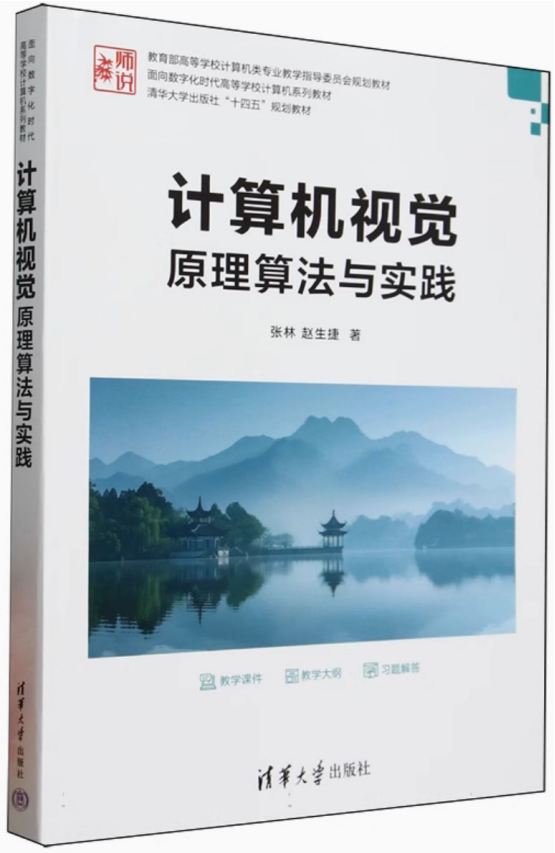
Our work related to ML

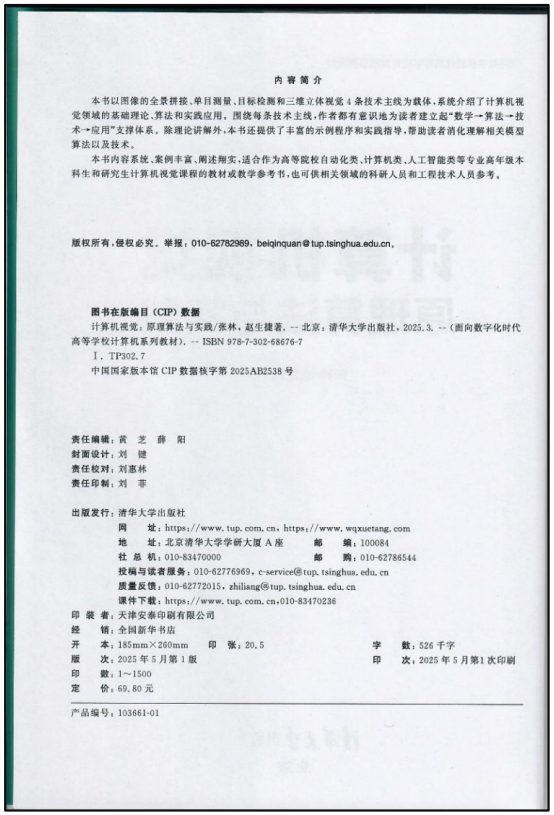
Speech-Driven Zero-Shot System for 6D Robotic Arm Grasping



Yitong Ge, Lin Zhang*, Yang Chen, and Ying Shen, WSGS: A Speech-Driven Zero-Shot System for 6D Robotic Arm Grasping, in: Proc. ICME, 2025.

A live demo for object detection





内容简介

本书以图像的全景拼接、单目测量、目标检测和三维立体视觉4条技术主线为载体，系统介绍了计算机视觉领域的基础理论、算法和实践应用。围绕每条技术主线，作者都有意识地为读者建立起“数学→算法→技术→应用”支撑体系。除理论讲解外，本书还提供了丰富的示例程序和实践指导，帮助读者消化理解相关模型算法以及技术。

本书内容系统、案例丰富、阐述翔实，适合作为高等院校自动化类、计算机类、人工智能类等专业高年级本科生和研究生计算机视觉课程的教材或教学参考书，也可供相关领域的科研人员和工程技术人员参考。

图书在版编目（CIP）数据

计算机视觉：原理算法与实践/张林，赵生捷著. -- 北京：清华大学出版社，2025.3. --（面向数字化时代高等学校计算机系列教材）. -- ISBN 978-7-302-68676-7

Ⅰ. TP302.7

中国国家版本馆 CIP 数据核字第 2025AB2538 号

责任编辑：黄 芝 薛 阳
封面设计：刘 键
责任校对：刘惠林
责任印制：刘 菲

出版发行：清华大学出版社
网 址：https://www.tup.com.cn，https://www.wqxuetang.com
地 址：北京清华大学学研大厦A座 邮 编：100084
社 总 机：010-83470000 邮 购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：https://www.tup.com.cn，010-83470236

印 装 者：天津安泰印刷有限公司
经 销：全国新华书店
开 本：185mm×260mm 印 张：20.5 字 数：526千字
版 次：2025年5月第1版 印 次：2025年5月第1次印刷
印 数：1～1500
定 价：69.80元

产品编号：103661-01

- ✓ 教育部高等学校计算机类专业教学指导委员会规划教材
- ✓ 面向数字化时代高等学校计算机系列教材
- ✓ 清华大学出版社“十四五”规划教材

张林，赵生捷，《计算机视觉：原理算法与实践》，清华大学出版社，2025年5月

上课理念与教材编写指导思想

CV教学中的三个问题

CV知识点分散，如何将它们有机组织在统一的框架之下？

如何培养学生解决“未来问题”的能力？也就是如何培养学生分析问题、逻辑推理、形成方案、迭代优化的综合能力？

如何在这门课上提升学生综合应用各门类数学知识的能力？

理念

以解决某个具体技术问题为载体，将相关知识点循序渐进地组织在一起，建立起“数学→算法→技术→应用”的逻辑支撑体系，强调“CV课程本身也只是个载体，以此为载体提升数学知识的运用能力、应对问题的逻辑分析能力，是更加重要的”

本书特点

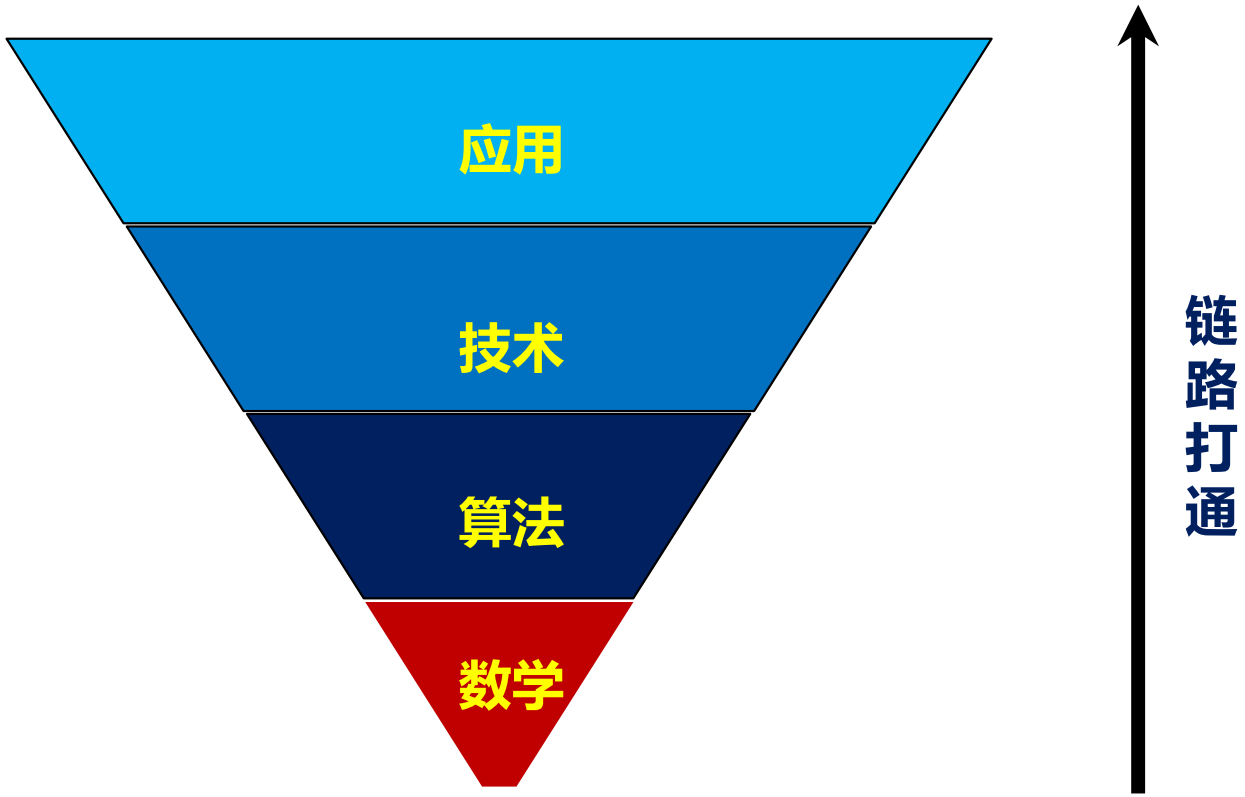
需要提升的学生能力

- 解决问题的逻辑分析能力
- 数学知识运用能力
- 将算法转换为实现的编程能力

本书（及配套材料）内容

- 按4条技术主线，构建“数学→算法→技术→应用”逻辑支撑体系
- 注重数学过程的理论推导
- 强调实践，提供与理论内容完全对应的全套示范代码与操作指南

本书特点

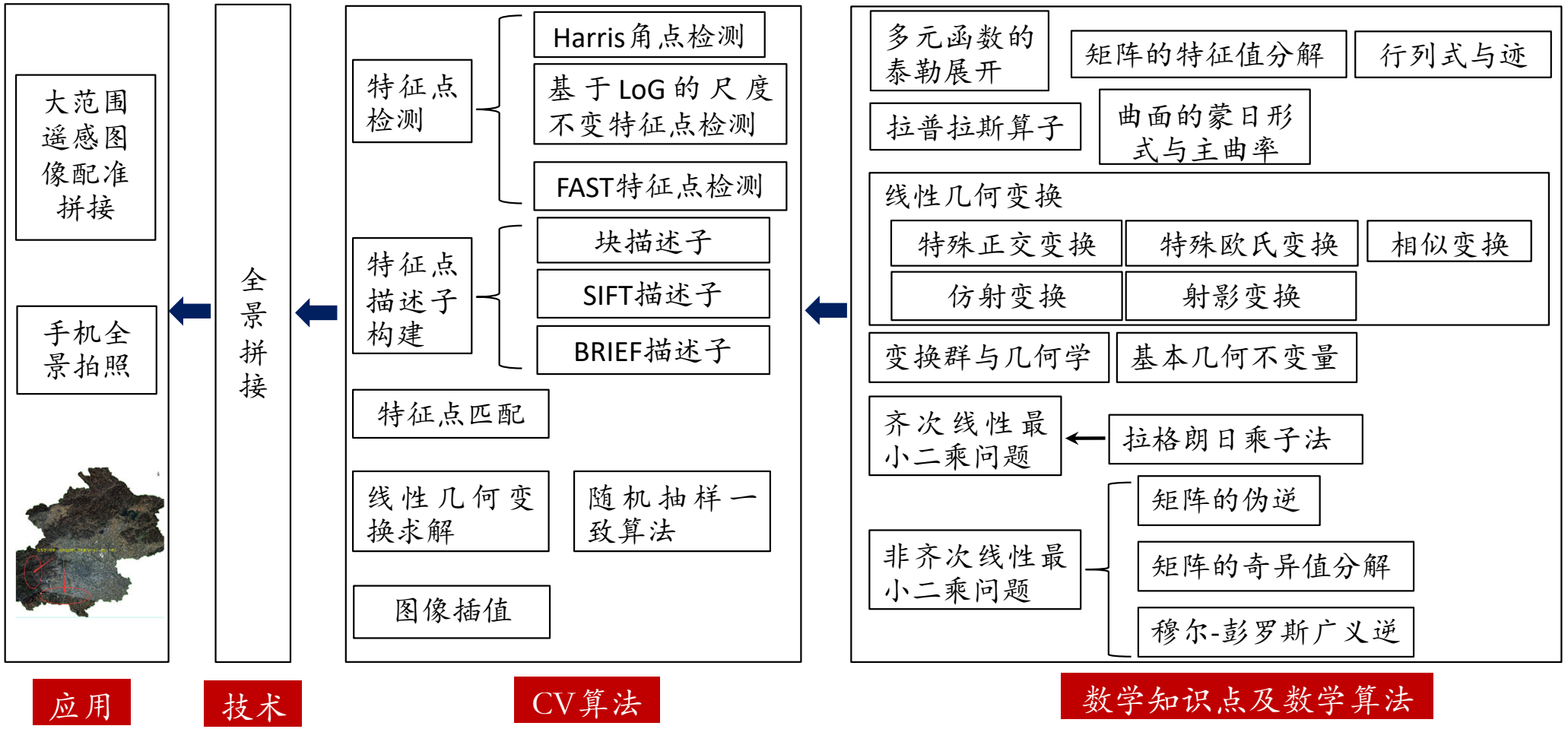


本书及配套材料内容

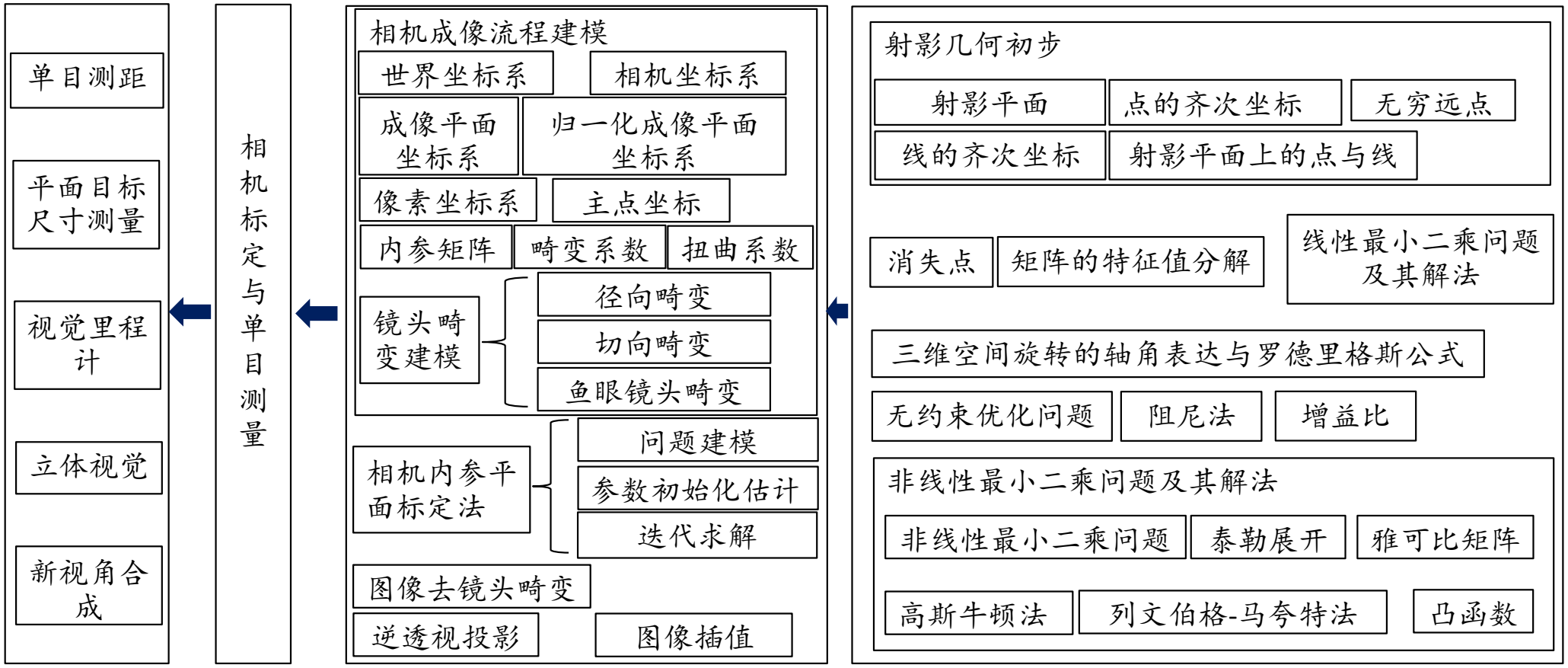
本书配套代码、教学PPT、勘误表、教学视频请见

https://github.com/csLinZhang/CVBook

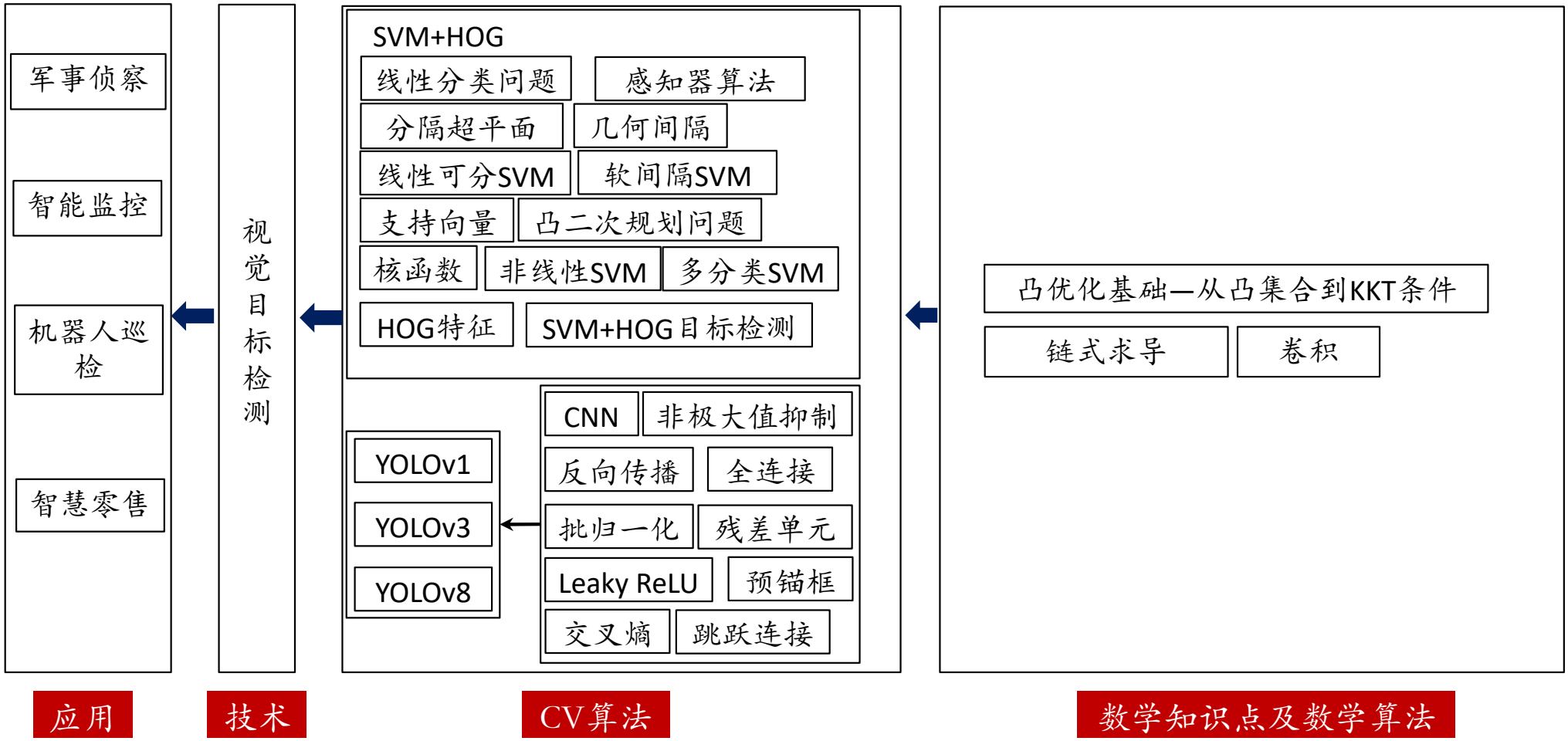
按四条技术主线，有机组织知识点—全景拼接



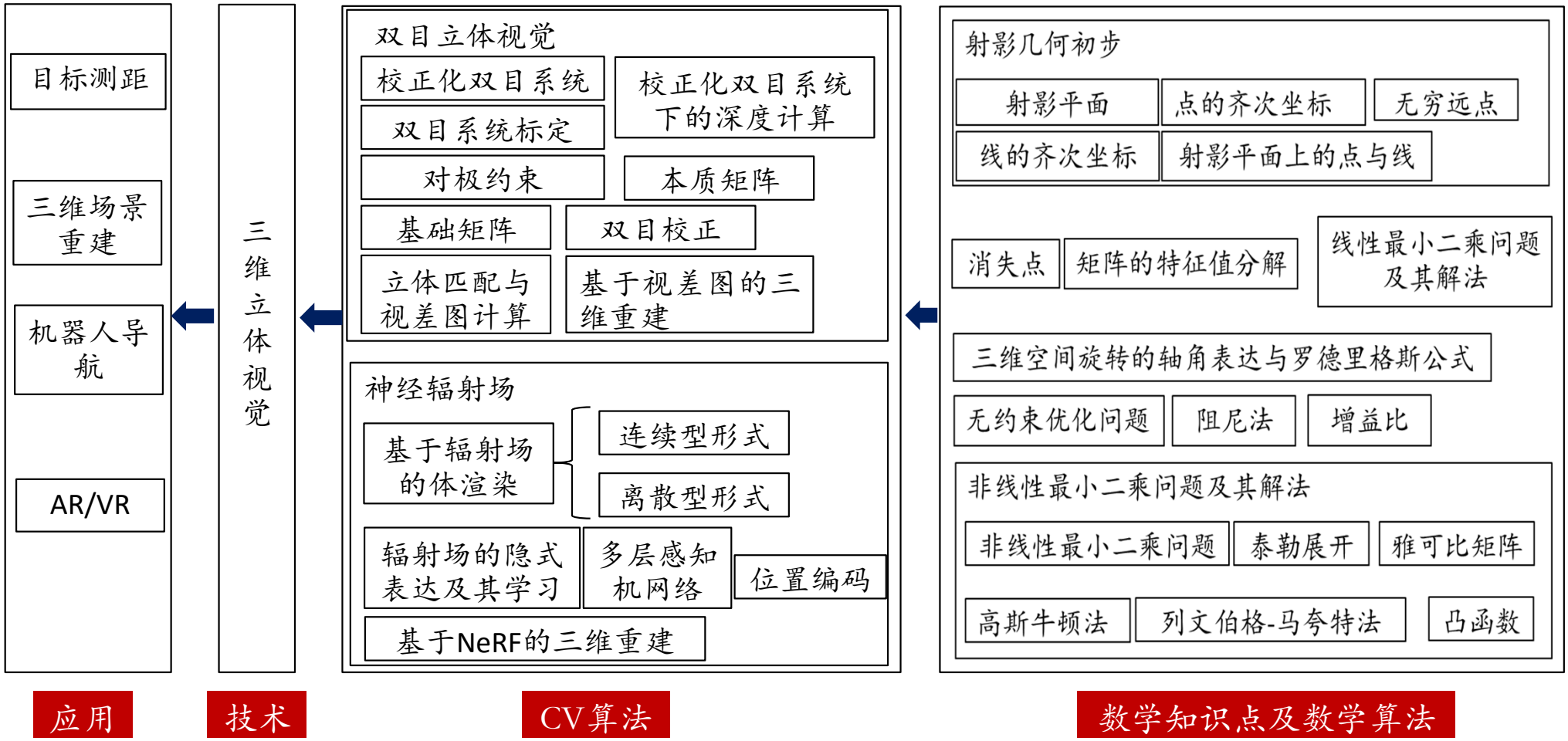
按四条技术主线，有机组织知识点—单目测量



按四条技术主线，有机组织知识点—目标检测



按四条技术主线，有机组织知识点—三维立体视觉



注重实践，提供全部代码与操作指南

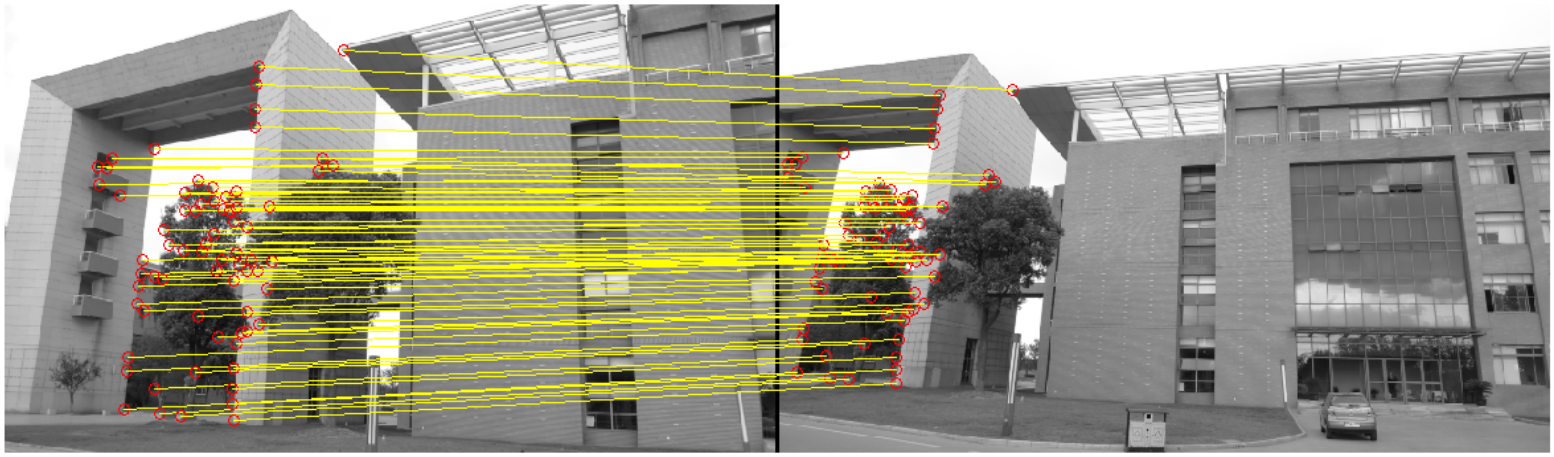
提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

全景拼接

注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

全景拼接



注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

全景拼接

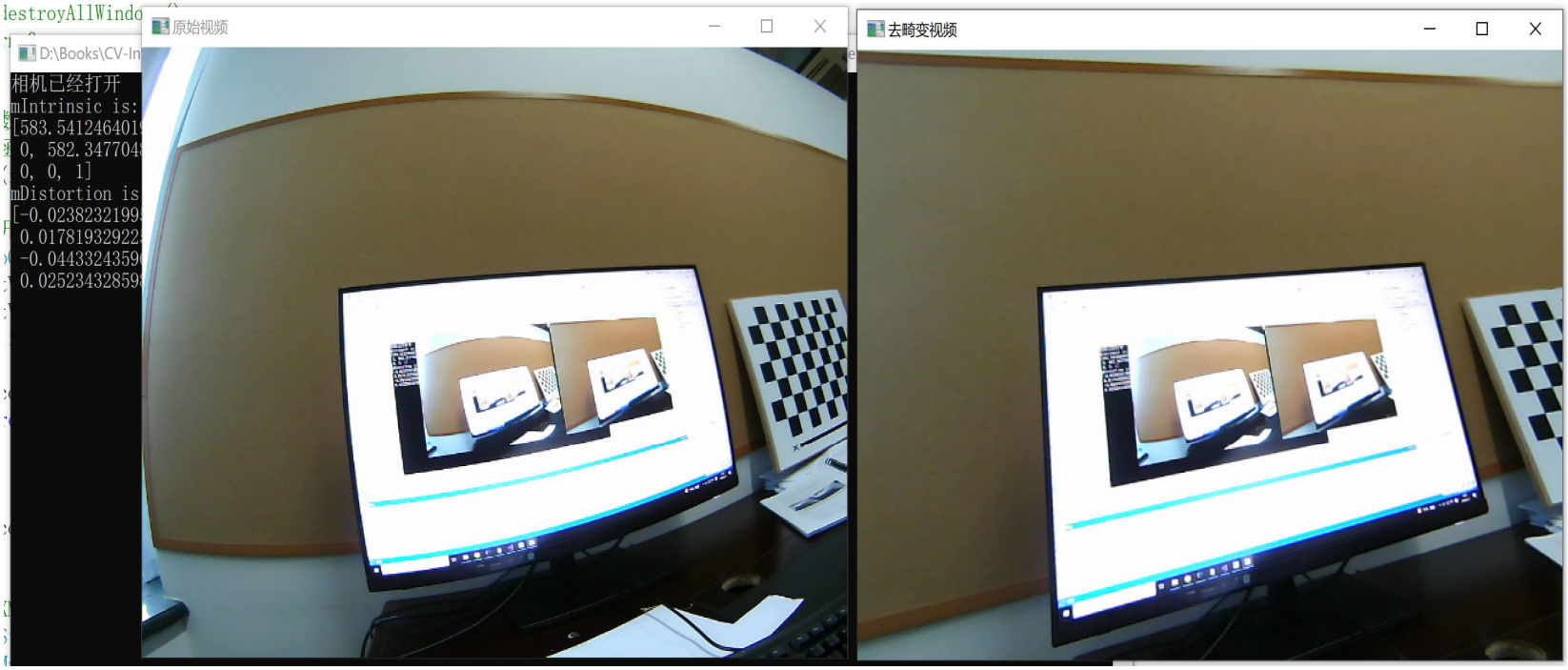
注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

全景拼接

注重实践，提供全部代码与操作指南

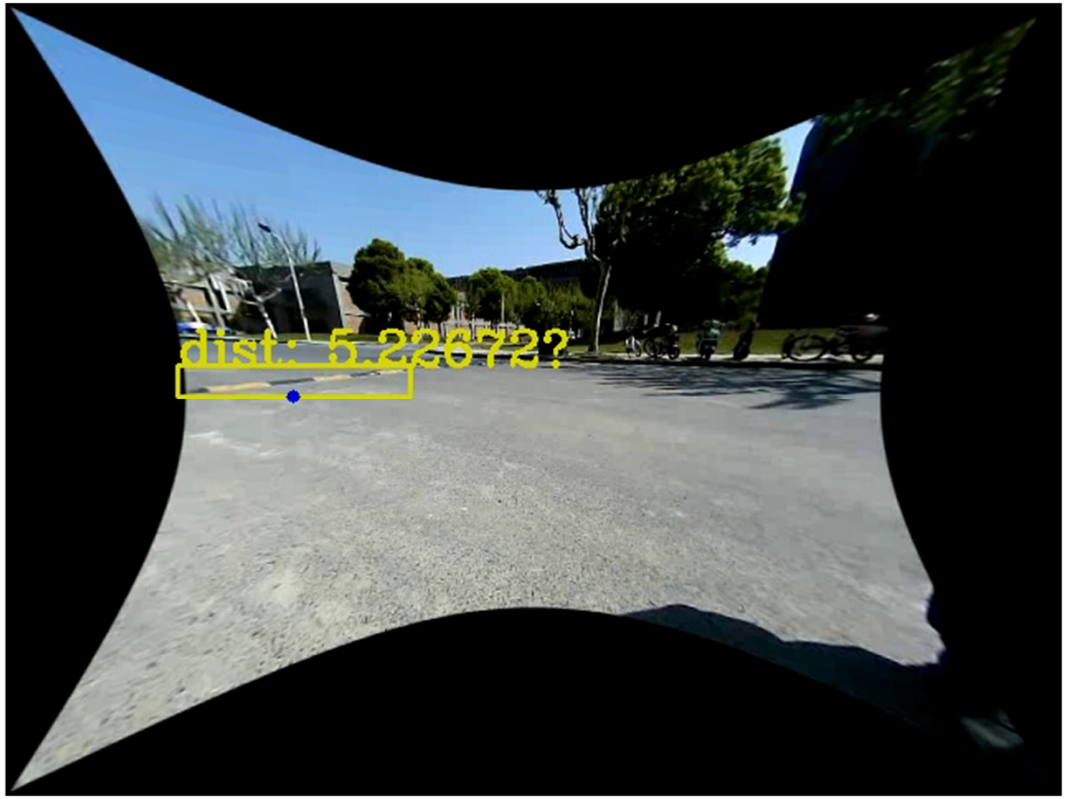
提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径



相机标定与单目测量

注重实践，提供全部代码与操作指南

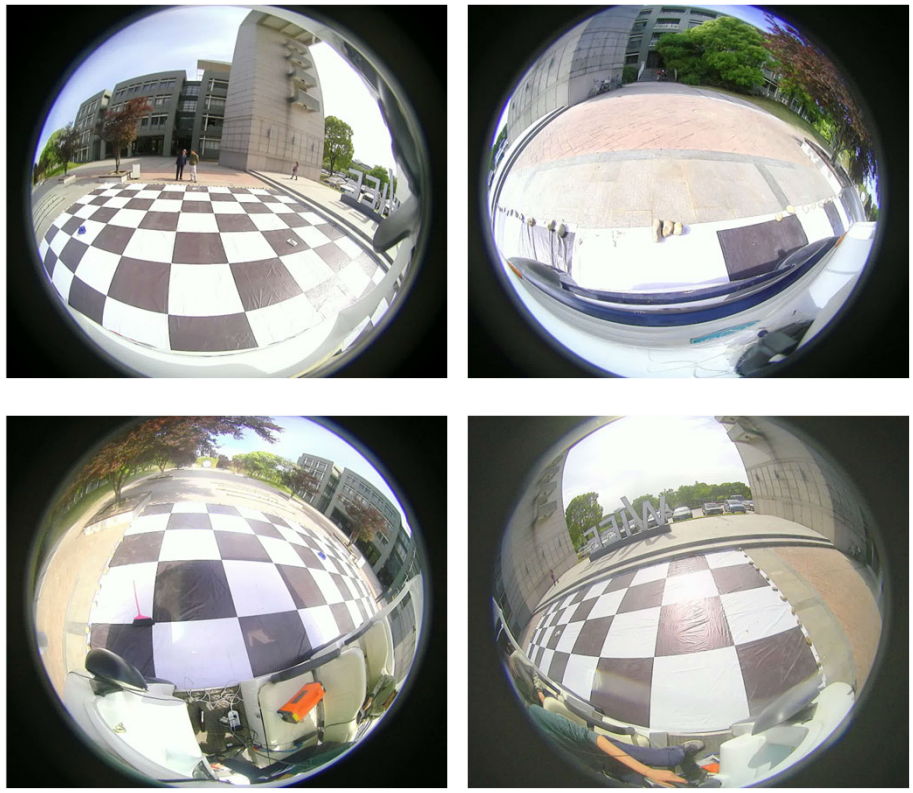
提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径



相机标定与单目测量

注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径



相机标定与单目测量

注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

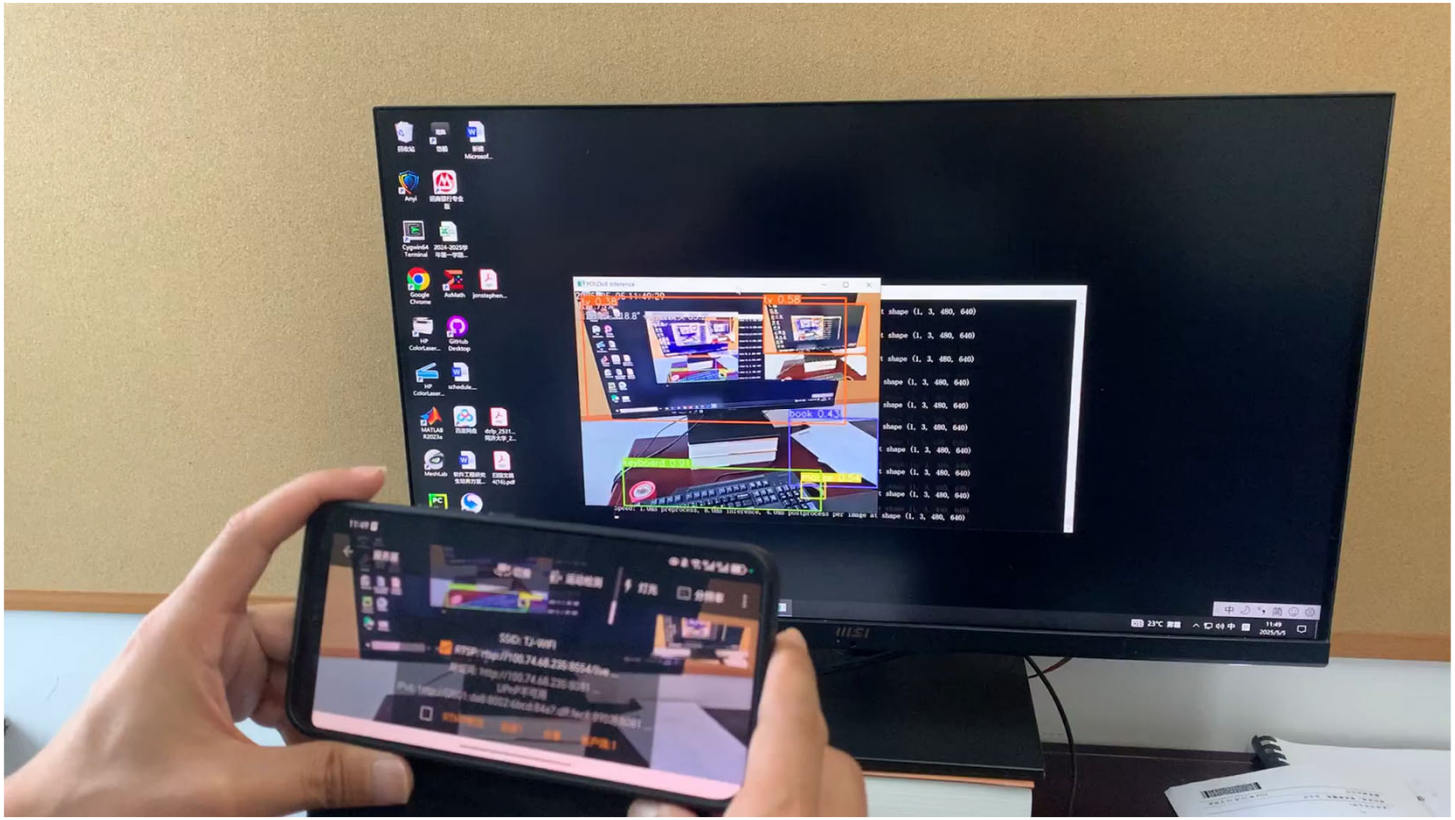
目标检测

行人与减速带检测

注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

目标检测



手机摄像头作为IP Camera

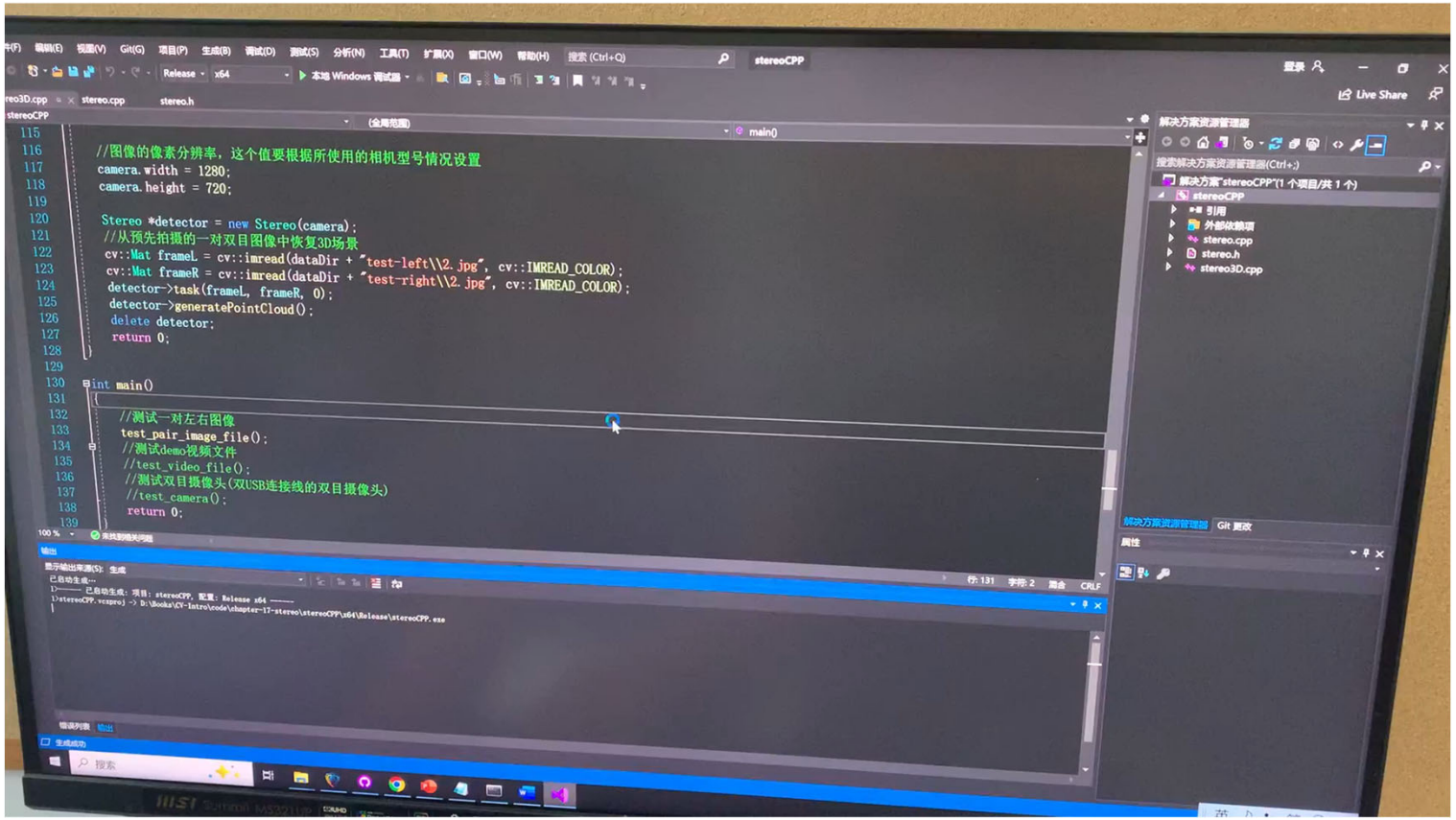
注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

三维立体视觉

注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

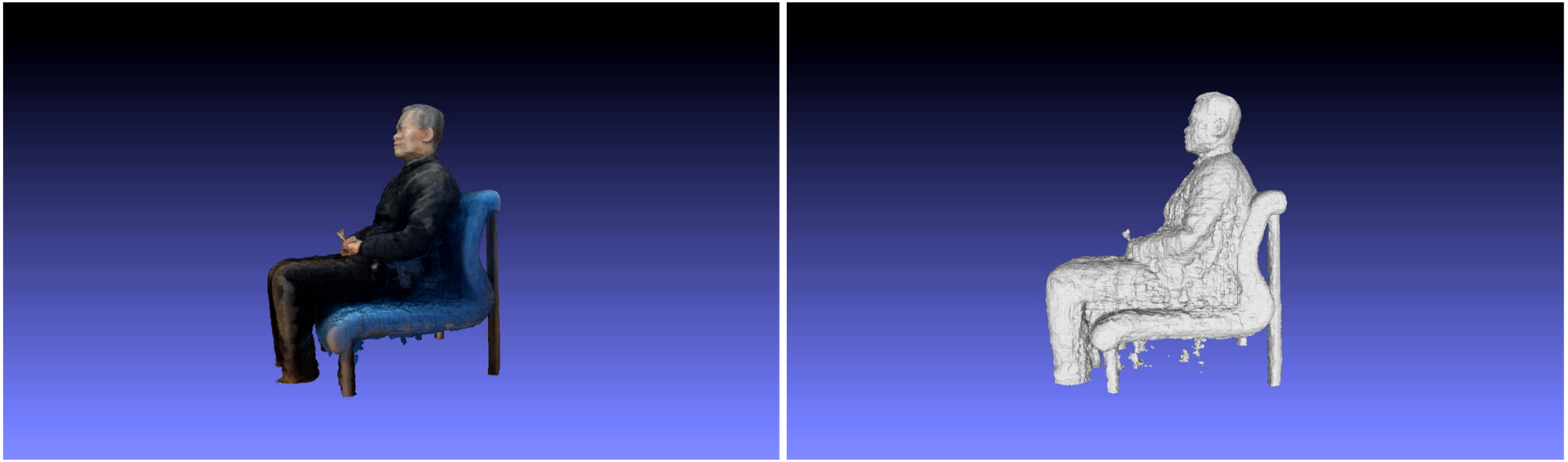


三维立体视觉

注重实践，提供全部代码与操作指南

提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径

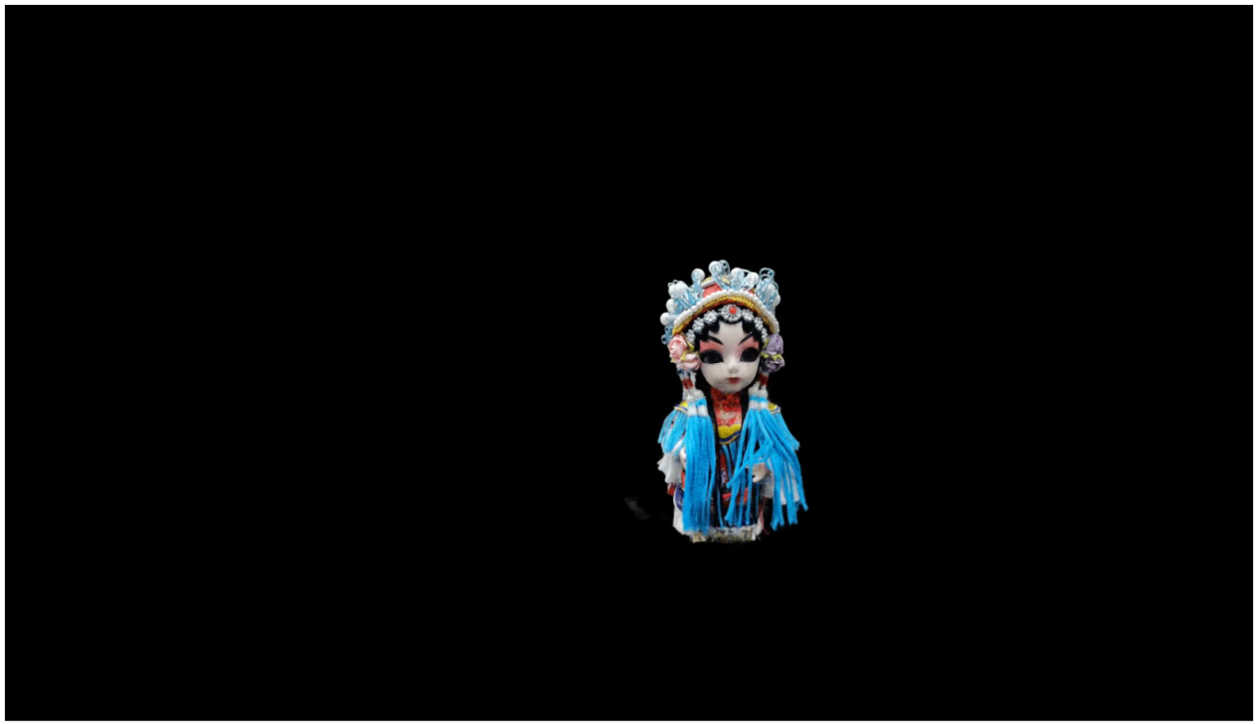
三维立体视觉



基于NeRF的三维重建

注重实践，提供全部代码与操作指南

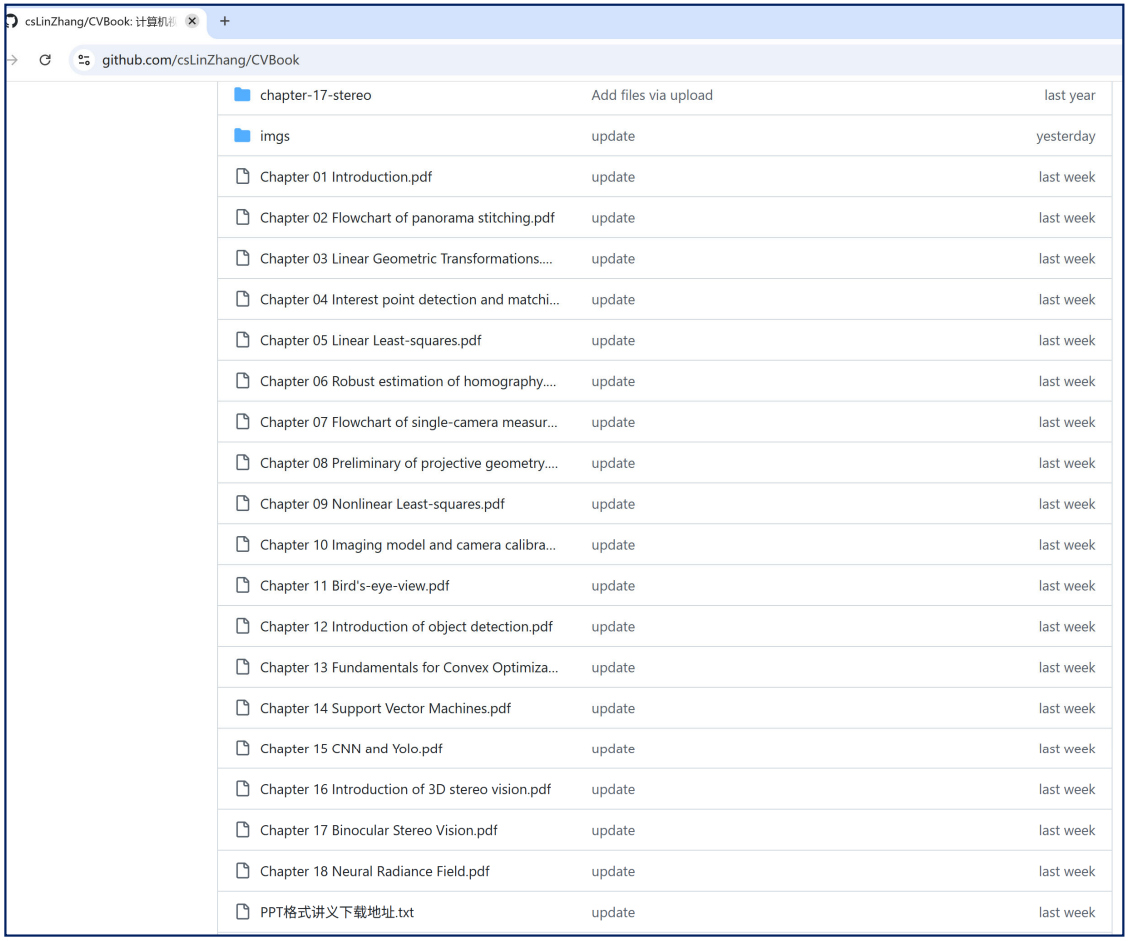
提供全部源代码或操作指南，协助学生打通“数学→算法→技术→应用”学习路径



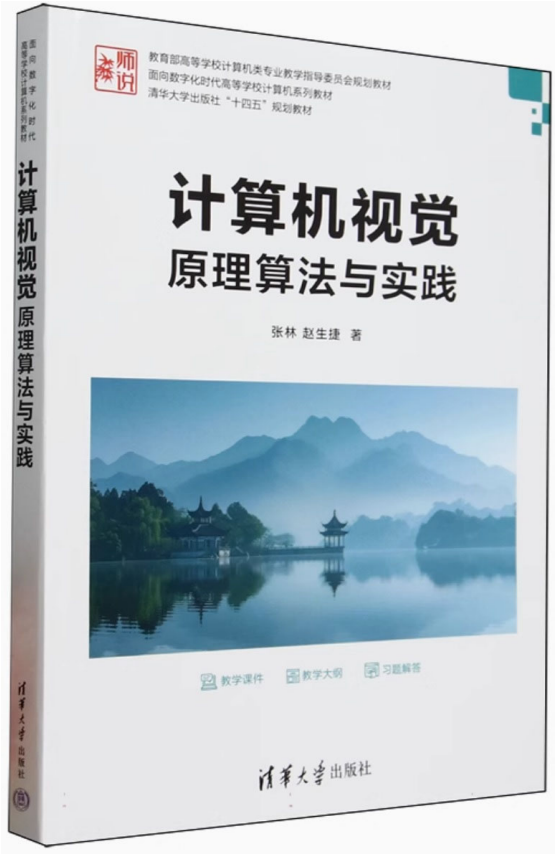
Instant-NGP

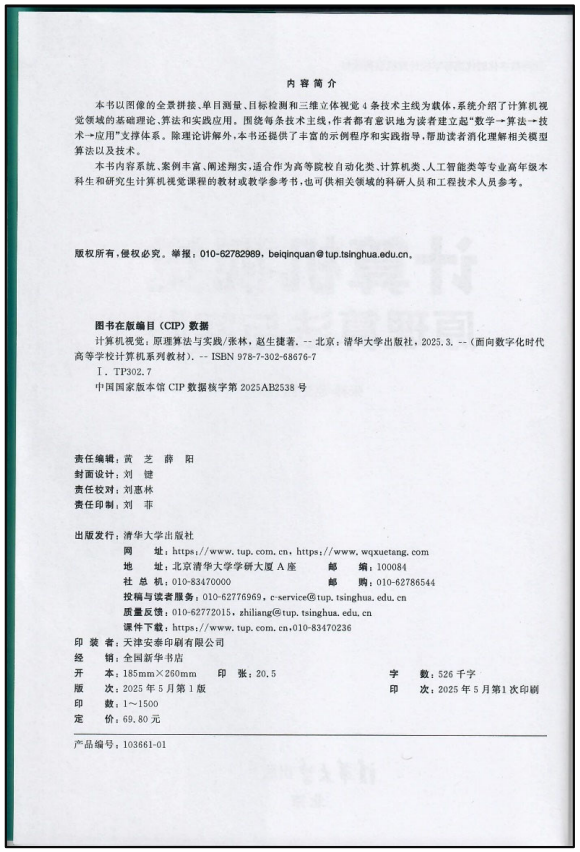
三维立体视觉

提供全套授课PPT格式课件



chapter-17-stereo	Add files via upload	last year
imgs	update	yesterday
Chapter 01 Introduction.pdf	update	last week
Chapter 02 Flowchart of panorama stitching.pdf	update	last week
Chapter 03 Linear Geometric Transformations....	update	last week
Chapter 04 Interest point detection and matchi...	update	last week
Chapter 05 Linear Least-squares.pdf	update	last week
Chapter 06 Robust estimation of homography....	update	last week
Chapter 07 Flowchart of single-camera measur...	update	last week
Chapter 08 Preliminary of projective geometry....	update	last week
Chapter 09 Nonlinear Least-squares.pdf	update	last week
Chapter 10 Imaging model and camera calibra...	update	last week
Chapter 11 Bird's-eye-view.pdf	update	last week
Chapter 12 Introduction of object detection.pdf	update	last week
Chapter 13 Fundamentals for Convex Optimiza...	update	last week
Chapter 14 Support Vector Machines.pdf	update	last week
Chapter 15 CNN and Yolo.pdf	update	last week
Chapter 16 Introduction of 3D stereo vision.pdf	update	last week
Chapter 17 Binocular Stereo Vision.pdf	update	last week
Chapter 18 Neural Radiance Field.pdf	update	last week
PPT格式讲义下载地址.txt	update	last week





内容简介

本书以图像的全景拼接、单目测量、目标检测和三维立体视觉4条技术主线为载体，系统介绍了计算机视觉领域的基础理论、算法和实践应用，围绕每条技术主线，作者都有意识地为读者建立起“数学→算法→技术→应用”支撑体系。除理论讲解外，本书还提供了丰富的示例程序和实践指导，帮助读者消化理解相关模型算法以及技术。

本书内容系统、案例丰富、阐述翔实，适合作为高等院校自动化类、计算机类、人工智能类等专业高年级本科生和研究生计算机视觉课程的教材或教学参考书，也可供相关领域的科研人员和工程技术人员参考。

图书在版编目（CIP）数据

计算机视觉：原理算法与实践/张林，赵生捷著. -- 北京：清华大学出版社，2025.3. --（面向数字化时代高等学校计算机系列教材）. -- ISBN 978-7-302-68676-7

Ⅰ. TP302.7

中国国家版本馆CIP数据核字第2025AB2538号

责任编辑：黄 芝 薛 阳
封面设计：刘 键
责任校对：刘惠林
责任印制：刘 菲

出版发行：清华大学出版社
网 址：https://www.tup.com.cn，https://www.wqxuetang.com
地 址：北京清华大学学研大厦A座 邮 编：100084
社 总 机：010-83470000 邮 购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：https://www.tup.com.cn，010-83470236

印 装 者：天津安泰印刷有限公司
经 销：全国新华书店
开 本：185mm×260mm 印 张：20.5 字 数：526千字
版 次：2025年5月第1版 印 次：2025年5月第1次印刷
印 数：1～1500
定 价：69.80元

产品编号：103861-01

欢迎选用本教材，谢谢大家☺

傍晚，小街路面上沁出微雨后的湿润，和煦的细风吹来，抬头看看天边的晚霞，嗯，明天又是一个好天气。走到水果摊旁，挑了个根蒂蜷缩、敲起来声音浊响的青绿西瓜，一边满心期待着皮薄肉厚瓤甜的爽落感，一边愉快地想着：这学期狠下了功夫，基础概念弄得清清楚楚，算法作业也是信手拈来，这门课成绩一定差不了！

摘自《机器学习》（周志华著，2016）

What is machine learning?

- Gives "computers the ability to learn without being explicitly programmed" (Arthur Samuel in 1959)

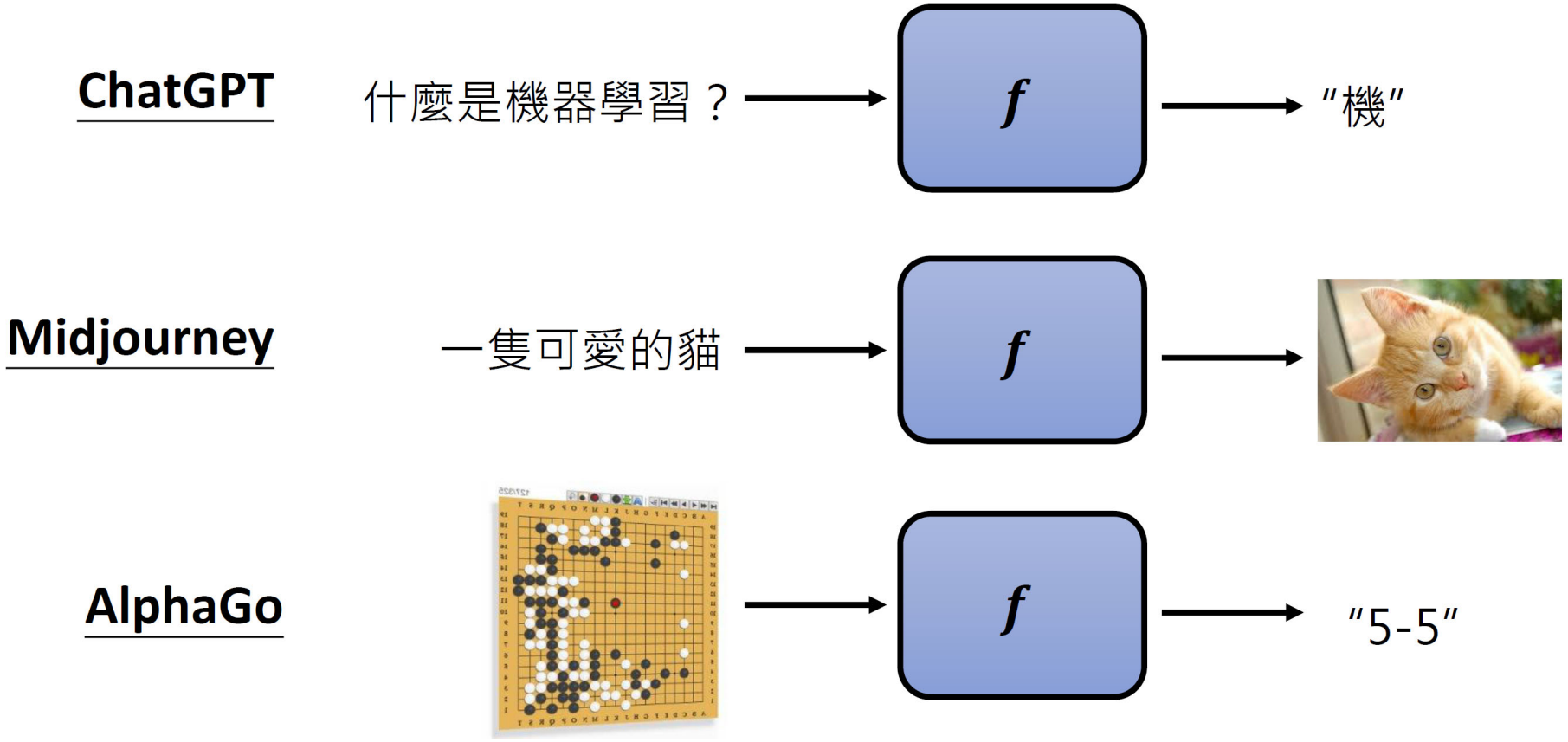
Arthur Lee Samuel
(December 5, 1901 – July 29, 1990)

- It explores the study and construction of algorithms that can learn from and make predictions on data
- It is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or unfeasible

[1] Samuel, Arthur L., Some Studies in Machine Learning Using the Game of Checkers, IBM Journal of Research and Development, 1959

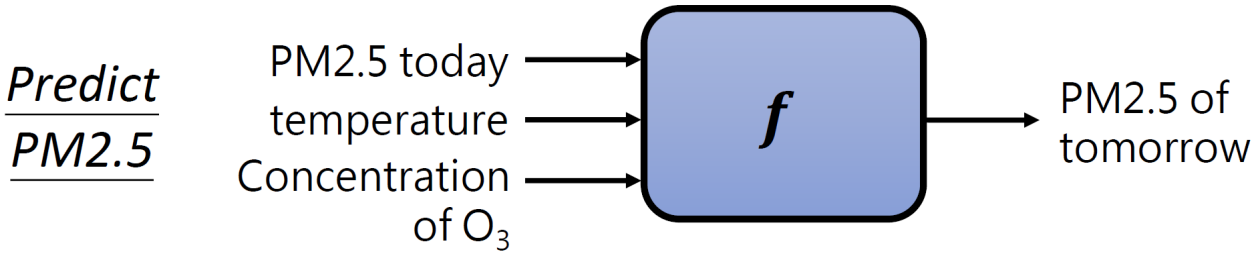
What is machine learning?

- The core goal of machine learning is to find a function f

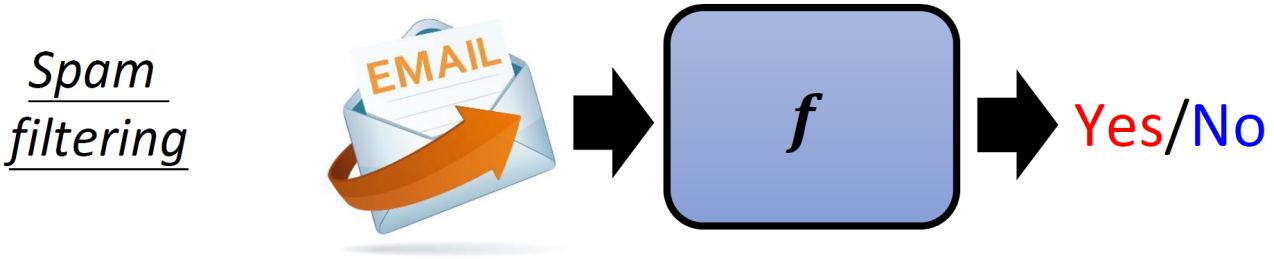


What is machine learning?

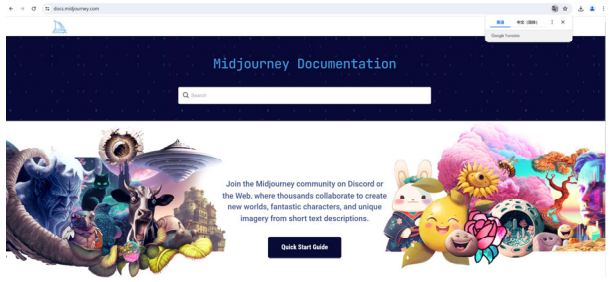
- According to the output, the machine learning problems can be categories as classification, regression and generative learning
  - Regression: the output of the function is a number



  - Classification: the output of the function is a choice from a fixed set



  - Generative learning: the output is a structured object (e.g., an image, a sentence)



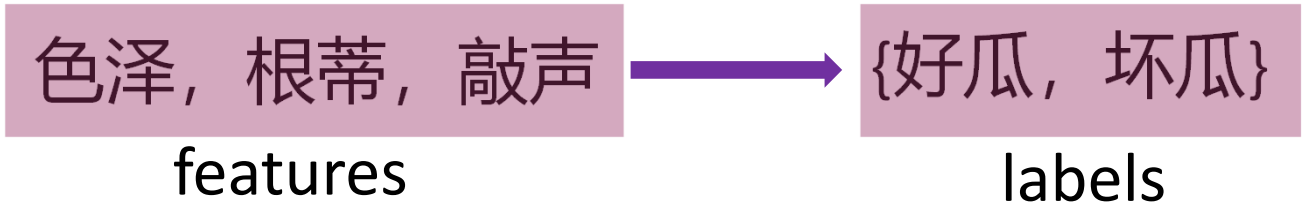
What is machine learning?

- Three steps to identify f

Determine the function set	Specify the set of the functions	Model
	Deep learning (CNN, RNN, Transformer), SVM (hyperplanes), Decision tree, etc.	
Determine the criteria	How to judge whether a function is “good” or “bad”?	Loss
	Supervised learning, Semi-supervised learning, Reinforcement learning, etc.	
Figure out the “best” function	According to the criteria, how to identify the ‘best’ function from the set?	optimization
	Gradient descent, convex optimization, generic algorithm, etc.	

About sample

- Attribute (feature), attribute value, label, and example



{青绿，蜷缩，浊响：好瓜}

feature values　label value

one example

Training, testing, and validation

- Training sample and training set

  A training set comprising m training samples,

$$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$$

  where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id}) \in \boldsymbol{\chi}$ is the feature vector of ith sample and $y_i \in \boldsymbol{\zeta}$ is its label

  By training, our aim is to find a mapping,

$$f : \boldsymbol{\chi} \mapsto \boldsymbol{\zeta}$$

  based on D

  If $\boldsymbol{\zeta}$ comprises discrete values, such a prediction task is called "**classification**"; if it comprises real numbers, such a prediction task is called "**regression**"

Training, testing, and validation

- Training sample and training set
- Test set
  - A test set is a set of data that is independent of the training data, but that follows the same probability distribution as the training data
  - Used only to assess the performance of a fully specified classifier

Training, testing, and validation

- Training sample and training set
- Test set
- Validation set
  - In order to avoid overfitting, when any classification parameter needs to be adjusted, it is necessary to have a validation set; it is used for model selection
  - The training set is used to train the candidate algorithms, while the validation set is used to compare their performances and decide which one to take

Overfitting, Generalization, and Capacity

- Overfitting
  - It occurs when a statistical model describes random error or noise instead of the underlying relationship
  - It generally occurs when a model is excessively complex, such as having too many parameters relative to the number of observations
  - A model that has been overfit will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data

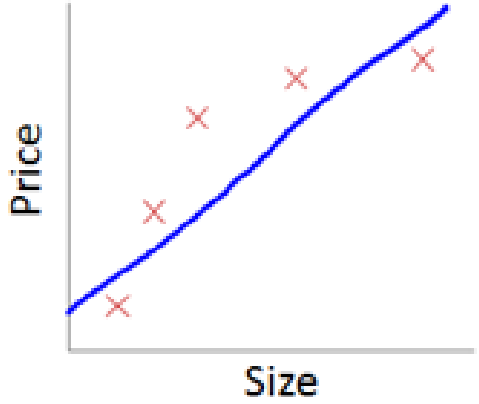
Overfitting, Generalization, and Capacity

- Overfitting
- Generalization
  - Refers to the performance of the learned model on new, previously unseen examples, such as the test set

Overfitting, Generalization, and Capacity

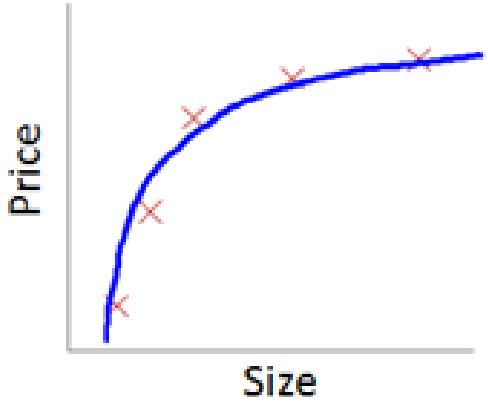
- Overfitting
- Generalization

Example: Linear regression (housing prices)



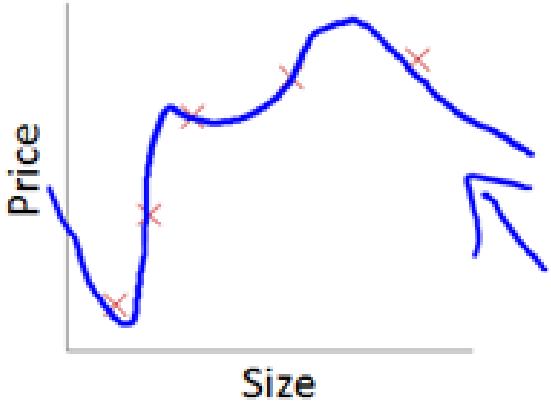
$\rightarrow \theta_0 + \theta_1 x$

"Underfit" "High bias"



$\rightarrow \theta_0 + \theta_1 x + \theta_2 x^2$

"Just right"



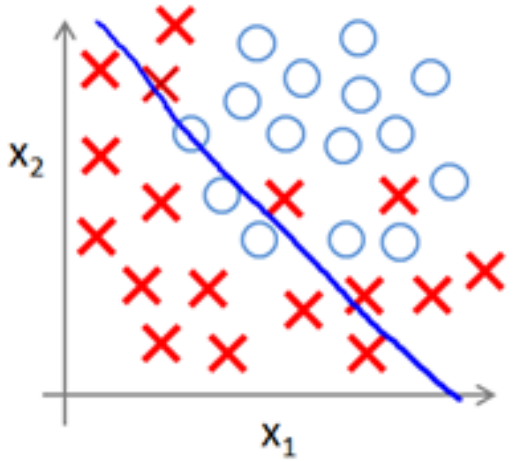
$\rightarrow \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

"Overfit" "High variance"

Overfitting, Generalization, and Capacity

- Overfitting
- Generalization

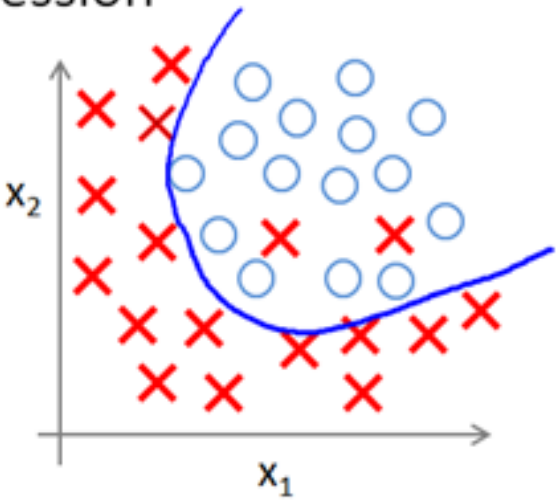
Example: Logistic regression



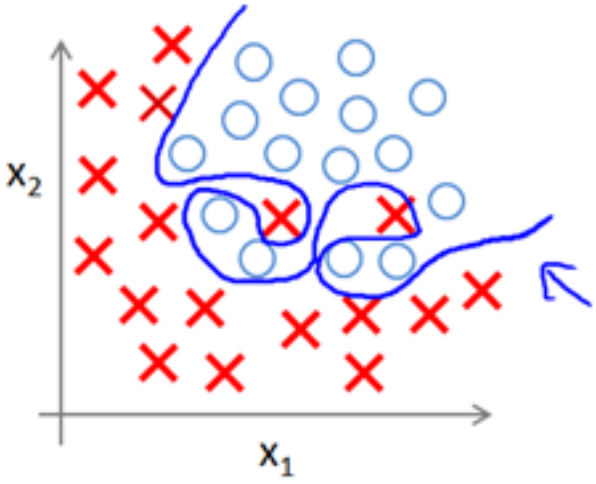
$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

(g = sigmoid function)

"Underfit"



$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$



$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \dots)$

"Overfit"

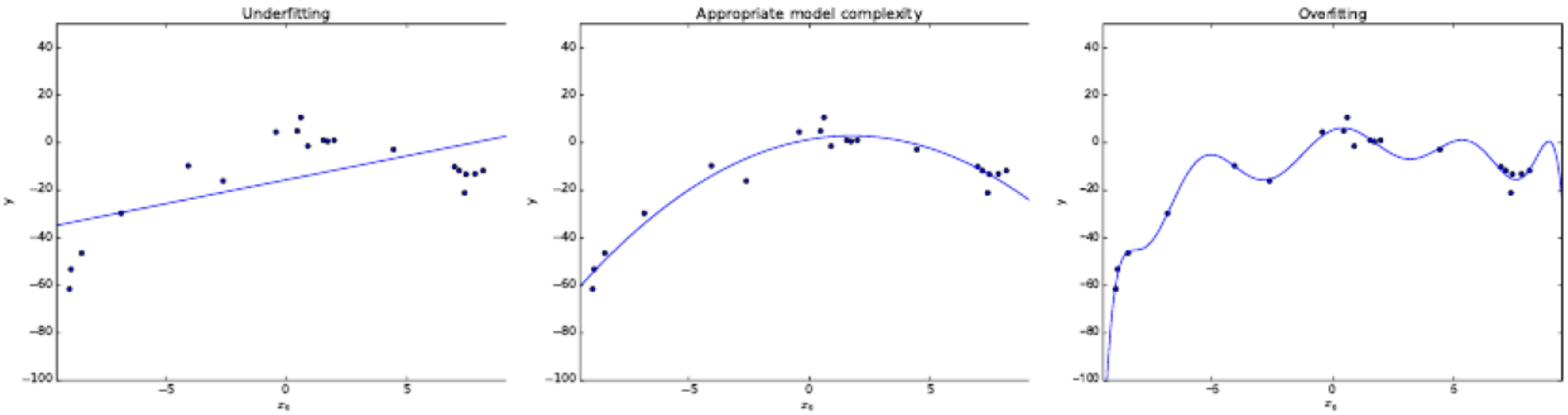
Overfitting, Generalization, and Capacity

- Overfitting
- Generalization
- Capacity
  - Measures the complexity, expressive power, richness, or flexibility of a classification algorithm
  - Ex, DCNN (deep convolutional neural networks) is powerful since its capacity is very large

$$y^* = b + \omega x, \quad y^* = b + \omega_1 x_1 + \omega_2 x_2, \quad y^* = b + \sum_{i=1}^{10} \omega_i x_i$$

higher capacity

Overfitting, Generalization, and Capacity



higher capacity

Performance Evaluation

Given a sample set (training, validation, or test)

$$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$$

To assess the performance of the learner f, we need to compare the prediction $f(\boldsymbol{x})$ and its ground-truth label y

For regression task, the most common performance measure is MSE (mean squared error),

$$E(f; D) = \frac{1}{m}\sum_{i=1}^{m}(f(\boldsymbol{x}_i) - y_i)^2$$

Performance Evaluation (for classification)

- Error rate
  - The ratio of the number of misclassified samples to the total number of samples

$$E(f;D)=\frac{1}{m}\sum_{i=1}^{m}\mathbf{1}\left(f(\boldsymbol{x}_i)\neq y_i\right)$$

- Accuracy
  - It is derived from the error rate

$$acc(f;D)=\frac{1}{m}\sum_{i=1}^{m}\mathbf{1}\left(f(\boldsymbol{x}_i)=y_i\right)=1-E(f;D)$$

Performance Evaluation (for classification)

- Precision and Recall

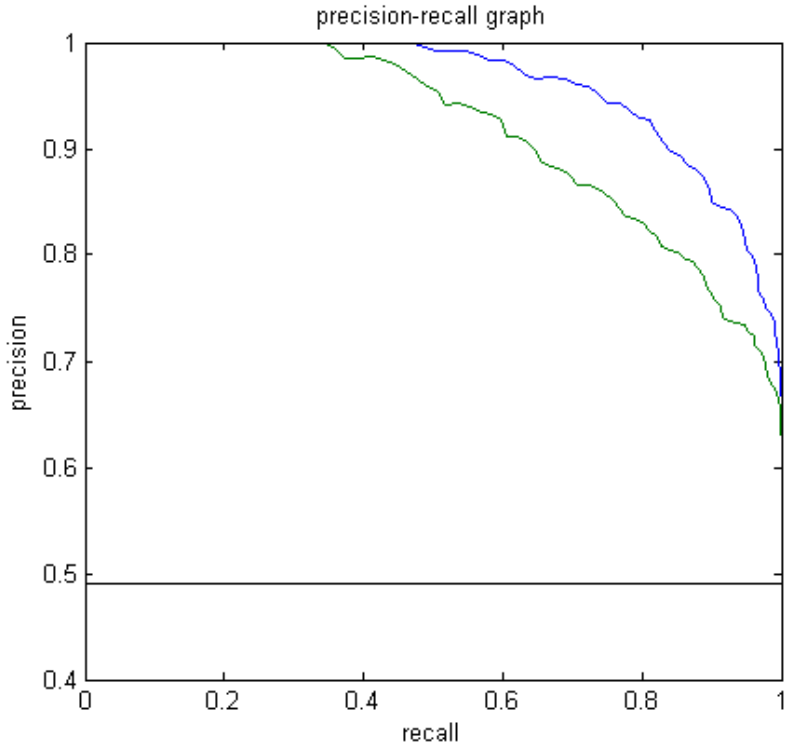
Ground truth	Prediction	
	positive	negative
positive	True Positive (TP)	False Negative (FN)
negative	False Positive (FP)	True Negative (TN)

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

Performance Evaluation (for classification)

- Precision and Recall
  - Often, there is an inverse relationship between precision and recall, where it is possible to increase one at the cost of reducing the other
  - Usually, PR-curve is not monotonic



Performance Evaluation (for classification)

- Precision-recall should be used together; it is meaningless to use only one of them
- However, in many cases, people want to know explicitly which algorithm is better; we can use F-measure

$$F_\beta = \frac{(1+\beta^2)\times P \times R}{(\beta^2 \times P) + R}$$

Performance Evaluation (for classification)

- To derive a single performance measure

Varying threshold, we can have a series of (P, R) pairs,

$$(P_1, R_1), (P_2, R_2), \ldots, (P_n, R_n)$$

Then,

$$P_{macro} = \frac{1}{n}\sum_{i=1}^{n} P_i \qquad R_{macro} = \frac{1}{n}\sum_{i=1}^{n} R_i$$

$$F_{\beta-macro} = \frac{\left(1+\beta^2\right) \times P_{macro} \times R_{macro}}{\left(\beta^2 \times P_{macro}\right) + R_{macro}}$$

Model selection—Cross validation

- Simple cross validation
  - Split the dataset at hand into a training set and a validation set
  - Training the models on the training set, and selecting the best model based on the evaluation on the validation set
- S-fold cross validation
  - Randomly split the dataset at hand into S equal-sized subsets; any two subsets do not overlap with each other
  - For one learning model, train it on S-1 subsets and evaluate its performance on the remaining one subset; repeat such a training-evaluating procedure S times, each time using a different subset for evaluation; averaging the obtained S evaluation errors as the performance of this learning model
- Leave-one-out cross validation
  - It can be regarded as a special case of the S-fold cross validation strategy, i.e., $S=m$, m is the number of training samples

Class-imbalance Issue

- Problem definition
  - It is the problem in machine learning where the total number of a class of data is far less than the total number of another class of data
  - This problem is extremely common in practice
- Why is it a problem?
  - Most machine learning algorithms work best when the number of instances of each classes are roughly equal
  - When the number of instances of one class far exceeds the other, problems arise

Class-imbalance Issue

- How to deal with this issue?
  - Modify the cost function
  - Under-sampling, throwing out samples from majority classes
  - Oversampling, creating new virtual samples for minority classes
    - » Just duplicating the minority classes could lead the classifier to overfitting to a few examples
    - » Instead, use some algorithm for oversampling, such as SMOTE (synthetic minority over-sampling techniqe)[1]

[1] N.V. Chawla *et al*., SMOTE: Synthetic Minority Over-sampling Technique, J. Artificial Intelligence Research 16: 321-357, 2002

Class-imbalance Issue

- Minority oversampling by SMOTE[2]

Add new minority class instances by:

- For each minority class instance c
  - neighbours = Get KNN(5)
  - n = Random pick one from neighbours
  - Create a new minority class r instance using c's feature vector and the feature vector's difference of n and c multiplied by a random number
    - » i.e. r.feats = c.feats + (n.feats - c.feats) * rand(0,1)

[2] N.V. Chawla *et al*., SMOTE: Synthetic Minority Over-sampling Technique, J. Artificial Intelligence Research 16: 321-357, 2002


Q+A