
视觉与学习青年学者研讨会
VISION AND LEARNING SEMINAR
年度重要学术进展-8
上海交通大学
沈旭
大模型微调技术持续演进，大模型与小数据/小算力间的鸿沟不再难以跨越
VALSE2024相关活动：APR-9(基础大模型)、Tutorial-3(开放词汇)、Workshop-5(智慧医疗)/9(大模型迁移)
2022 | Stable Diffusion
2023.04 | SAM
2023.03 | GPT4
大模型
参数高效微调算法
LoRA (ICLR2022)
医学图像
单张 GPU 3090,2080, ...
小数据/小算力
QLoRA (NeurIPS2023)
SAM-COBOT (CVPR2024)
ControlNet (ICCV2023 best paper)
LISA (1.4K stars in 8 months)
自然语言处理领域
计算机视觉领域
多模态领域

年度重要学术进展-6

谢丽彤 德莱德大学

视觉与学习青年学者研讨会
VISION AND LEARNING SEMINAR

SAM出现并获得广泛应用，X Anything范式开始流行

VALSE2024相关活动：APR-9(基础大模型)、Tutorial-3(开放词汇)、Workshop-9(大模型迁移)

SAM被迅速应用至各种垂直域

- 医学影像分割
- 点云分割
- 遥感图像分割
- 全景图像分割
- 高效数据标注
- 视频超分辨率
- 视频目标追踪
- 机器人模型X
- 开放词汇目标检测
- 3D场景重建/理解
- 目标姿态估计
- 图像编辑
- 字幕生成
- 风格迁移
- 目标计数
- ……





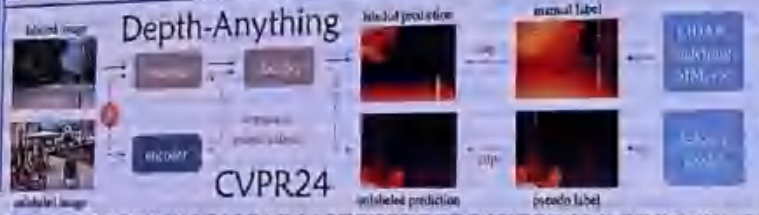

SAM（Segment Anything Model）
ICCV 2023 最佳论文提名

改善SAM效率、稳定性和可扩展性

- BLO-SAM, ICML24
- Conv-LoRA, ICLR24
- MobileSAM
- Finetune-SAM
- ……

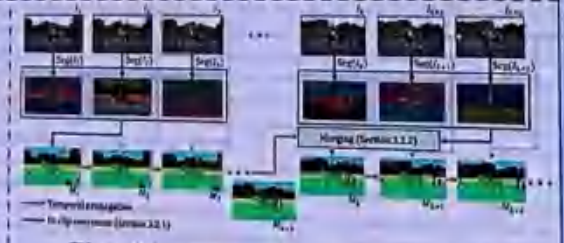
EfficientSAM

FastSAM



- Inpaint-Anything
- Matting-Anything
- Caption-Anything
- Anything-3D
- Ask-Anything
- ……

SAM成功推动了X Anything
研究范式的发展



Tracking Anything, ICCV23

年度重要学术进展-5

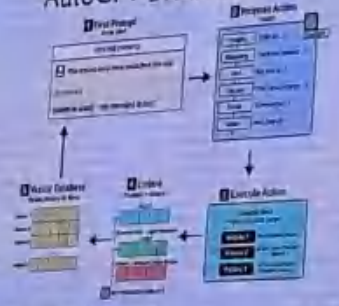
视觉与学习青年学者研讨会
VISION AND LEARNING SEMINAR

上海交通大学
陈思衡

以AutoGPT为代表的AI agent范式开始流行，带来AI生产效率提升

VALSE2024相关活动：APR-2(智能体视觉)、Tutorial-1(具身智能与智能体)

AutoGPT 引发持续跟进



AI小镇火遍全球，agent持续出圈

Aloha 2家务机器人效果拔群

2023年3月 AutoGPT github开源，Google发布Palm-E具身多模态模型

2023年4月 斯坦福AI小镇发表"Generative agents: Interactive simulacra of human behavior"

2023年6月 MetaGPT框架正式开源，截至目前收获近4w star

2023年7月 CMU WebArena 框架发布，WebAgent迎来大发展

2023年11月 OpenAI上线GPT Shop低代码平台GPT Builder，大大降低agent搭建难度

2024年1月 李飞飞团队发布综述："Agent AI: Surveying the Horizons of Multimodal Interaction"

2024年2月 微软上线Windows agent框架UFO，两个月时间收获4K stars；斯坦福团队发布Aloha 2家务机器人，引发全网热议

从纯语言到**多模态**到**具身智能**，从"做一件事"到"让AI做一件事"：学术界/产业界/创投界正以空前的热情投入AI agent的研究与发展

年度重要学术进展-4

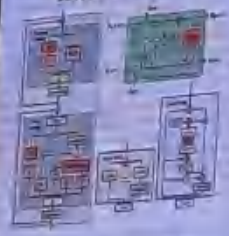
华中科技大学
王兴刚

视觉与学习青年学者研讨会
VISION AND LEARNING SEMINAR

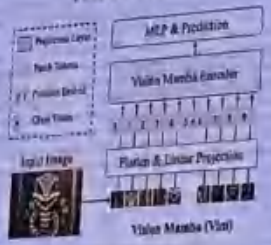
高效大语言模型持续发展，新型CV计算架构持续涌现

VALSE2024相关活动：APR-8(新型高效网络架构)/9(基础大模型)、Workshop-12(大模型理论与机理)

RWKV Network

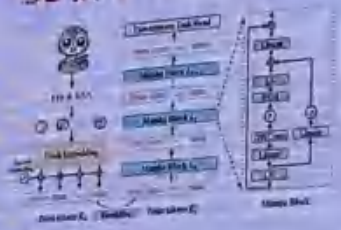


Vision Mamba

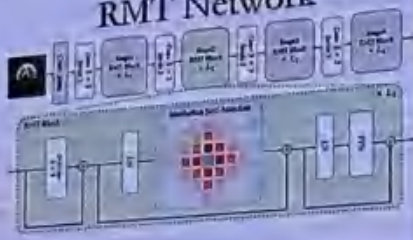


U-Mamba

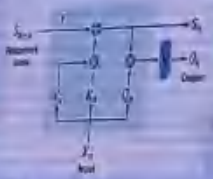
3D点云 PointMamba



RMT Network



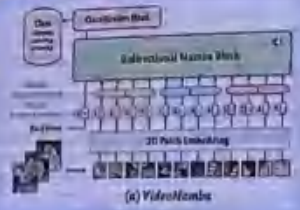
Retention Network



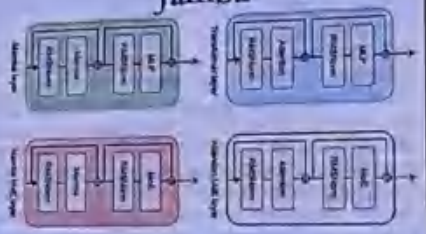
VMamba

VM-UNet

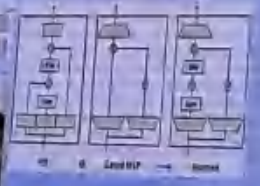
视频理解 VideoMamba



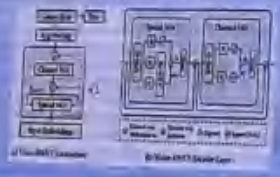
Jamba



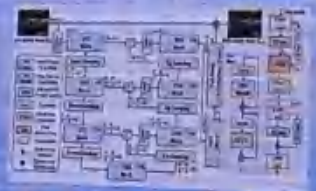
Mamba Network



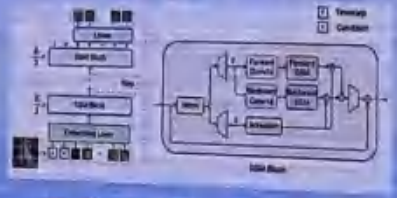
Vision RWKV



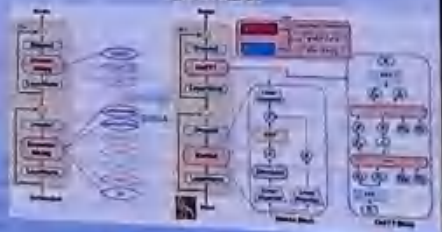
MambaIR / VmambaIR



扩散模型 DiS



Simba



高效大语言模型

骨干网络

医学图像&底层视觉

视觉理解&图像生成

混合架构&多模态

年度重要学术进展-3

北京大学 王鹤

视觉与学习青年学者研讨会 VISION AND LEARNING SEMINAR

具身视觉在大模型加持下显著扩展其范畴，CV算法与环境的交互愈发丰富

VALSE2024相关活动：APR-5(具身感知交互)/11(具身决策)、Tutorial-1(具身智能)、Workshop-4(具身智能)

大脑

PaLM-E - Google

GPT-4V - OpenAI

LLM- Planer [ICCV23]

EmbodiedGPT [NeurIPS23]

Voxposer [CoRL23 oral]

SAGE

小脑

AnyGrasp [TRO23]

RT-2 [CoRL23]

GR-1 [ICLR23]

RoboCook [CoRL23 best system paper]

Diffusion Policy [RSS 23]

Eurek

UniDexGrasp++ [ICCV23 best paper finalist]

NVM [CVPR 23 Highlight]

本体

LEAP Hand

Mobile Aloha

Hello Robot

Galbot

1X

Tesla Optimus

Digit

Fouri

Unitree H1

New Atlas

Figure 01

数据集

RoboGen

Open X-Embodiment

ManiSkill2

DexGraspNet [ICRA'23 outstanding manipulation finalist]

EmbodiedScan [CVPR24]

DROID

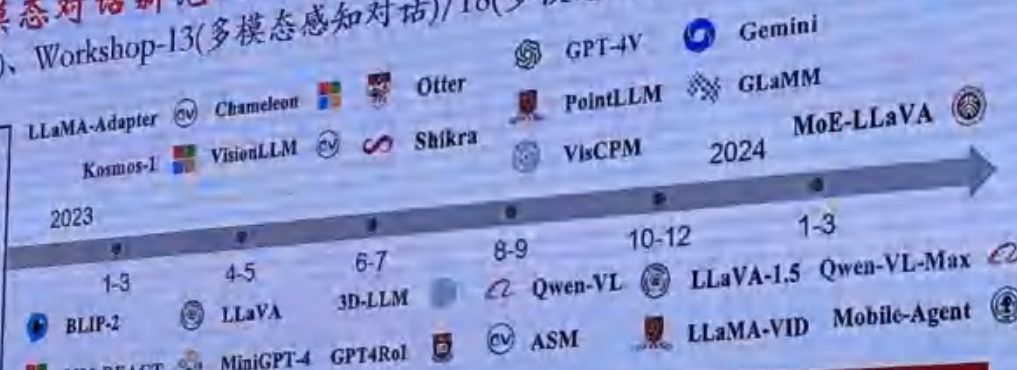
年度重要学术进展-2

鹏城实验室 杨文瀚

视觉与学习青年学者研讨会 VISION AND LEARNING SEMINAR

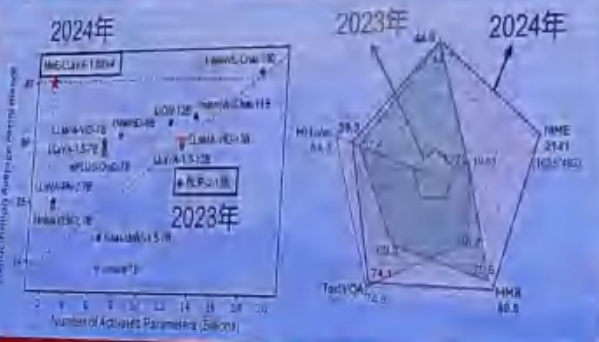
多模态大语言模型的进化，带来多模态对话新范式，扩展了视觉理解的外延

VALSE2024相关活动：APR-5(多模态感知交互)、Workshop-13(多模态感知对话)/18(多模态大模型)

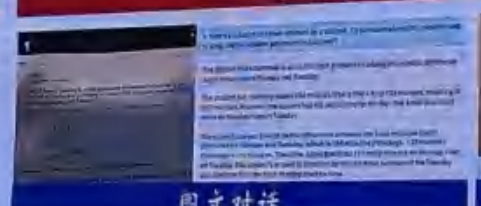


以大语言模型为抓手，实现跨模态对话问答

多模态对话系统百花齐放



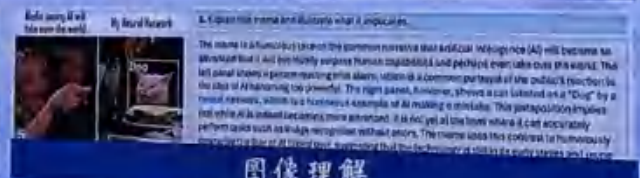
性能持续进步



图文对话

OCR

图像生成



图像理解

涌现出在图文对话、图像理解、OCR等方面的卓越能力

年度重要学术进展-1

视觉与学习青年学者研讨会 VISION AND LEARNING SEMINAR

北京大学 袁粒

高清图像和视频生成技术快速发展，基础模型渊现物理世界建模能力

VALSR2024相关活动：APR-4(视频生成)、Tutorial-2(视频生成)、Workshop-3(视频生成)

学术进展

2023.02-03

- OpenAI提出Consistency Models一步生成扩散模型
- ControlNet(斯坦福)和T2I-Adapter(北大)分别被提出，用于文生图精准控制

2023.04-06

- 港中文和上海AI Lab共同提出AnimateDiff文生动态GIF模型
- 德国马普所、MIT等提出点追踪算法的DragGAN模型

2023.07-10

- 华为基于Diffusion Transformer提出文生图模型PixArt-α
- 阿里提出图生视频Animate Anyone生成人物角色动画模型

2023.10-12

- Stability AI公开论文SD3，基于多模态DiT和Flow Model的文生图模型

2024.01-04

- 阿里达摩院提出EMO，图像和音乐生成人物画像视频模型
- 莫纳什大学和上海AI Lab等基于DiT提出2D+1D attention文生视频模型Latte

应用进展

2023.02-03

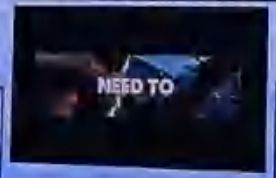
- Runway发布第一版商业文生视频应用Gen-1

2023.04-06

- Midjourney V5系列发布，文生图质量和美学新高度
- OpenAI推出文生图产品DALL-E3，与ChatGPT深度融合
- Stability AI正式公布最新的开源绘图应用——SDXL1.0

2023.07-10

- Pika发布文生视频工具Pika 1.0，Runway发布Gen-2



2023.10-12

- Stability AI开源Stable Video Diffusion

2024.01-04

- OpenAI放出Sora文生视频模型引发巨大关注，完全闭源仍无API接口和使用通道
- 生数科技推出Vidu但未开源，北大和NUS团队分别发起Sora复现计划

视觉与学习青年学者研讨会
VISION AND LEARNING SEMINAR

VALSE 2024

总结-变化总是悄然发生

CVPR2023所有论文（2359篇）标题的词云

CVPR2024所有论文（2719篇）标题的词云

不变的追求

- 视觉信号：Visual/Image/Video/3D
- 表征学习：Feature/Learning
- 高效计算：Model/Efficient

消逝的过往

- 传统任务：Object Detection
- 传统数据：Point Cloud
- 模型崇拜：Transformer

兴起的未来

- 结合语言：Language Model
- 生成任务：Generation
- 生成模型：Diffusion Model