

# 王晓东

✉ wangxiaodong21s@stu.pku.edu.cn · ☎ 155-7972-1557 · 🏠 主页

## 🎓 教育背景

北京大学, 软件与微电子学院 2021.9 – 2024.7  
硕士 (推荐免试) 软件工程 (智能科技) GPA: 3.66/4.0 荣誉: 北京大学三好学生  
北京信息科技大学, 计算机学院 2017.9 – 2021.7  
学士 数据科学与大数据技术 GPA: 4.43/5.0, 排名: 1/32 荣誉: 国家奖学金, 校长奖学金

## 🔧 技能

- 语言: Python, C/C++, Java, HTML/CSS
- 工具: PyTorch, Linux, Git, LaTeX, Gradio, Transformers
- 英语: CET-6: 518, CET-4: 577

## 👨‍💻 实习经历

微软亚洲研究院 (MSRA). 自然语言计算组, NUWA 组, 获“明日之星”荣誉 2022.5 – 2023.11

### ► 多模态大模型的预训练和微调研究

- 多模态语言大模型的预训练: 独立完成大模型的预训练, 基于 Llama-13B 模型, 在文本-图像数据集预训练。在 MSCOCO 的图像描述上达到最新 SOTA, BLEU@4 为 68.7, CIDEr 为 191.9。
- 多任务多模态大模型高效微调: 负责设计和实现两阶段的对齐和指令微调。在 Vicuna-7B 中引入 Adapter, 在图像理解、编辑和生成任务上同时微调, 实现了理解用户意图的多任务多模态大模型。

### ► NUWA-3D: 探索扩散模型在 3D 摄影中的应用

- 以第一作者身份, 提出一种自监督扩散模型, 提出一种基于掩码增强的 UNet 作为模型主干。使模型只需在图片数据集上训练, 继承了文本生成图像预训练模型的知识, 实现了高质量的图像拓展, 效果超过 Stable-Diffusion 系列, 以及逼真、稳定的图像 3D 空间拓展。
- *Learning 3D Photography Videos via Self-supervised Diffusion on Single Images. IJCAI 2023 (CCF A).*

### ► NUWA-XL: 探索扩散模型在文本生成超长视频的应用

- 以核心成员身份, 设计了掩码时序扩散模型的主干网络, 设计了首尾帧预测中间多帧, 实现了掩码和已知帧的高效插入; 参与设计了“从粗到细”的 Diffusion over Diffusion 的结构学习视频分布, 使得模型可以生成长时序下内容依然连贯的、具有真实镜头切换的动画视频。
- *NUWA-XL: Diffusion over Diffusion for eXtremely Long Video Generation. ACL 2023 ORAL (CCF A).*

### ► Visual ChatGPT: 探索大语言模型如何对接一系列视觉模型, GitHub 已获得 3.4 万点赞

- 以核心成员身份, 负责整个项目的代码实现, 引入了中文版交互, 提出了模板 API, 负责 GitHub 的维护; 参与设计利用思维链 (CoT) 方式连接 ChatGPT 和多个视觉模型, 设计高效的提示词修饰视觉模型、用户请求和历史对话, 实现在聊天中接收、处理、生成图像, 自动调用多个视觉模型。
- *Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models.* (谷歌引用 200 余次)

旷视科技. Face++ 研究院, 人脸服务器组 2021.06 – 2021.09

### ► 探究对比学习在人脸识别的相关作用; 无监督域适应

- 调研人脸识别损失函数, 引入对比学习方法, 在人脸测试集 AgeDB30 和 CPLFW 上提升了性能。
- 以第一作者身份, 提出从平滑性的角度提升域适应性能: 强制弱和强的图像增强的预测一致性, 增强了实例判别性的学习; 提出一种新的不确定性指标, 在 4 个 benchmark 上涨点显著, 接近 SOTA。
- *Revisiting Unsupervised Domain Adaptation Models: a Smoothness Perspective. ACCV 2022 (CCF C).*

## 📁 项目经历

北京大学-广荣科技 联合课题. HealGPT 大模型健康咨询应用 (经费 5 万元) 2023.11 – 至今

北京大学. 导师实验室. “半监督源假设转移”. ICASSP 2024 (CCF-B) 2023.08 – 2023.09

自然语言计算组. “开放词汇的负责任视觉生成”. AAAI 2024 (CCF-A) 2023.06 – 2023.08

中科院计算所. 智能信息实验室, “域适应和迁移学习算法” 市优秀毕设 2020.12 – 2021.06