

尹相臣 (Xiangchen Yin)

☎ (+86) 18769941501
✉ yinxiangchen@mail.ustc.edu.cn
WeChat: yinxc_suxi123

研究领域

AIGC、扩散模型、图像/视频生成、底层视觉

教育背景

- 2023.09–至今 中国科学技术大学，计算机技术
硕士 硕士二年级在读
研究方向: 扩散模型和 AIGC。
- 2019.09–2023.06 曲阜师范大学，物联网工程
本科 以综合排名 1/50 (2%) 保研。

实习经历

- 2024.06-2024.11 理想汽车-智能空间团队
(线下三个月) AIGC 研究实习生 (导师校企合作项目)
研究方向: 姿态引导的人物图像生成; 数字人视频生成。
- 2023.10-2024.06 腾讯-优图实验室
计算机视觉应用研究实习生
研究方向: 视觉文本生成及编辑的文生图扩散模型研究。
- 2022.04–2023.04 清华大学汽车安全与节能国家重点实验室
大三下–大四下 科研实习生
研究方向: 夜晚场景下的多模态车道线感知算法研究。

论文发表

AAAI2025 **GRPose: Learning Graph Relations for Human Image Generation with Pose Priors**

Xiangchen Yin, Donglin Di, Lei Fan, ..., Xiao Sun, Xun Yang

Conference on Artificial Intelligence (AAAI) 已接收

第一作者, 人工智能 CCF A 类会议;

分数: Accept、Accept、Strong Accept (778);

<https://arxiv.org/abs/2408.16540>

简介: 提出一种基于图关联的 GRPose 扩散模型用于姿态引导人物图像生成。本文的方法基于 Adapter 架构进行姿态的空间控制, 通过构建图结构捕获不同的姿态部分的高级关联关系来提高姿态对齐能力。另外我们采用姿态感知损失进一步校正姿态。相比于 ControlNet 我们在姿态对齐的 AP 上提升了近 10 个点。

Submitted2024 **Structure-Guided Diffusion Transformer for Low-Light Image Enhancement**

Xiangchen Yin, Zhenda Yu, Longtao Jiang, Xun Yang, Xin Gao, Xiao Sun

IEEE Transactions on Multimedia (TMM) 一审大修。

第一作者, 中科院一区 Top, CCF B 类 & 清华 A 类。

简介: 提出一种基于结构引导的 Diffusion Transformer 框架, 用于低光照增强。模型效果非常好, 性能超过了 24 年为止的绝大多数 SOTA 模型。

TIV2024 **Adaptive Entropy Multi-modal Fusion for Nighttime Lane Segmentation**
Xinyu Zhang, **Xiangchen Yin***, Xin Gao, Tianheng Qiu, Li Wang, Jun Li
IEEE Transactions on Intelligent Vehicles
学生一作（导师一作），一作为清华大学张新钰研究员。无人驾驶中科院一区 Top.
<https://ieeexplore.ieee.org/document/10506546>
简介：一种基于多模态的熵融合车道线检测框架，解决夜晚的车道线感知问题。

Arxiv2023 **DEFormer: DCT-driven Enhancement Transformer for Low-light Image and Dark Vision**
Xiangchen Yin, Zhenda Yu, Xin Gao, Ran Ju, Xiao Sun
Arxiv, 2023
第一作者，ICASSP2025（CCF B 类会议）在投，分数：Accept、Bordline Accept、bordline Reject；

ICANN2023 **PE-YOLO: Pyramid enhancement network for dark object detection**
Xiangchen Yin, Zhenda Yu, Zetao Fei, Wenjun Lv, Xin Gao
International Conference on Artificial Neural Networks, 2023 (ICANN '23)
第一作者，CCF C，谷歌学术 29 引用。
https://link.springer.com/chapter/10.1007/978-3-031-44195-0_14
简介：一种金字塔增强网络的预处理器，用于夜间检测。

项目

视觉文本编辑 银行卡场景下的视觉文本编辑应用
来自腾讯-优图实验室实习期间应用项目
平台：应用验证在 112 张 Telsa V100 上进行；
简介：在当前的 OCR 网络训练下，文字数据集的准确性和标注人力耗费巨大，通过 AIGC 模型来实现银行卡 OCR 识别网络所需要的数据集合成。基于 ControlNet 架构，我们实现了银行卡文本编辑模型的优化。
主要实现如下：（1）基于 ControlNet 实现银行卡文本编辑模型，给定银行卡号作为 Prompt 和被 mask 掉的银行卡图片为 Condition，实现银行卡的数据集图片合成，这种方式自带标签；（2）针对位置不对的问题，我们采用银行卡单字符训练标签和预训练的 UNet 分割网络校正位置；（3）针对字形不准确的问题，我们采用预训练的 OCR 编码器来校正特征。

数字人生成 实时性的音频驱动数字人视频生成系统
理想汽车上车项目
简介：输入一段语音和参考图片，可以生成一段根据语音唇形的数字人视频。
主要实现如下：（1）输入一段语音（或文字）和参考图像，基于 MuseTalker 来实现音频驱动的视频生成，EdgeTTS 将输入文字转为语音；（2）系统支持实时语音识别和视频字幕，使得用户可以通过语音与数字人进行自然的交流；（3）采用理想自研的 MindGPT 实现了数字人的多轮对话功能，能够理解并维持上下文相关的连贯对话；（4）在数字人视频生成后采用超分辨率模型提高了视频整理的分辨率。

获奖情况

2022 第十五届中国大学生计算机设计大赛全国总决赛三等奖
2021 第十二届蓝桥杯全国软件信息技术人才大赛全国总决赛二等奖（全国 2.5%）