

ZHICHENG ZHANG

☎ 185-6589-8403 · ✉ zhichengzhang@mail.nankai.edu.cn · 🌐 nku-zhichengzhang

EDUCATION

Nankai University *PhD*, CS, **Advisor:** Prof. Jufeng Yang 2021.9 - now
Xidian University *Bachelor*, EE 2017.9 - 2021.6

RESEARCH (MULTIMODAL LLM/VIDEO GENERATION & UNDERSTANDING)

[Submit to NeurIPS25] VidEmo: Causal Affective-Cues Reasoning for Emotion-centric VideoLLM
Zhicheng Zhang, Weicheng Wang, Yongjie Zhu, Wenyu Qin, Pengfei Wan, Di Zhang, Jufeng Yang **First Author**

[ICML25 Spotlight] MODA: MODular Duplex Attention for Multimodal Perception, Cognition, and Emotion Understanding
Zhicheng Zhang, Wuyou Xia, Chenxi Zhao, Zhou Yan, Xiaoqiang Liu, Yongjie Zhu, Wenyu Qin, Pengfei Wan, Di Zhang, Jufeng Yang [\[PDF\]](#) [\[Project\]](#) **First Author**

[CVPR24] ExtDM: Distribution Extrapolation Diffusion Model for Video Prediction
Zhicheng Zhang, Junyao Hu, Wentao Cheng, Danda Paudel, Jufeng Yang [\[PDF\]](#) [\[Code\]](#) [\[Project\]](#) **First Author**

[CVPR24] MART: Masked Affective RepresenTation Learning via Masked Temporal Distribution Distillation
Zhicheng Zhang, Pancheng Zhao, Eunil Park, Jufeng Yang [\[PDF\]](#) [\[Code\]](#) [\[Project\]](#) **First Author**

[CVPR23] Weakly Supervised Video Emotion Detection and Prediction via Cross-Modal Temporal Erasing Network
Zhicheng Zhang, Lijuan Wang, Jufeng Yang [\[PDF\]](#) [\[Code\]](#) [\[Video\]](#) **First Author**

[ICCV23] Multiple Planar Object Tracking
Zhicheng Zhang, Shengzhe Liu, Jufeng Yang [\[PDF\]](#) [\[Code\]](#) [\[Project\]](#) [\[MPOT-3K Dataset\]](#) [\[Video\]](#) [\[Demo\]](#) **First Author**

[TNNLS24] PlaneSeg: Building a Plug-in for Boosting Planar Region Segmentation
Zhicheng Zhang, Song Chen, Zichuan Wang, Jufeng Yang [\[PDF\]](#) [\[Code\]](#) **First Author**

[ACM MM22] Temporal Sentiment Localization: Listen and look in Untrimmed Videos
Zhicheng Zhang, Jufeng Yang [\[PDF\]](#) [\[Code\]](#) [\[TSL300 Dataset\]](#) [\[Video\]](#) **First Author**

[CVPR24] LAKE-RED: Camouflaged Images Generation by Latent Background Knowledge Retrieval-Augmented Diffusion
Pancheng Zhao, Peng Xu, Pengda Qin, Deng-ping Fan, *Zhicheng Zhang*, et al. [\[PDF\]](#) [\[Code\]](#) **Second Student Author**

INTERNSHIP

Kuaishou Kling AI 2024.9-2025.5

- **Main Responsibility.** Addressed the problem of **Visual-Text modality bias in large-scale interactive models for digital humans**. (1) Visual-Text Modality Bias: Existing multimodal large models exhibit attention bias toward the text modality, often neglecting critical visual information and discarding 39% of important visual cues. (2) Models and Algorithms: a. Dual-stream self-modal and cross-modal attention mechanisms based on attention decomposition. b. Multimodal large models tailored for perception, cognition, and emotion. (3) Outcomes: Accepted one first-author paper at **ICML2025 (Spotlight, 2.6% acceptance rate)**.
- **Main Responsibility.** Solved the technical challenge of **fine-grained video caption (aka. Video2Text)**. (1) Data: a. Collected 2,000 hours of portrait video data, constructing QA datasets and high-quality fine-grained caption data focused on facial attributes and emotional actions. b. Developed a temporal segmentation mechanism based on expression-action alignment to structure the data. (2) Models and Algorithms: a. Inference mechanism based on a portrait attribute process reward model. b. Caption models based on Qwen2.5-VL and LLaVA-NeXT-OV. (3) Outcomes: Submitted one first-author paper to **NeurIPS2025**; technical applications supported the Kling2.0 semantic response.

R&D PROJECTS FROM INDUSTRY

HUAWEI All-Weather Video Semantic Understanding 2021.9-2022.8

- **Project Leader** Addressed technical challenges in **obstacle detection and debris tracking** with high precision and lightweight models in security scenarios. (1) Data: Real-world data collected from high-mounted high-way surveillance cameras, including time-period filtering, object debris simulation, low-light enhancement for nighttime, and perspective simulation enhancement. (2) Models and Algorithms: Real-time debris tracking based on YOLOv5 and ByteTrack. (3) Deployment: Ported the solution to Huawei's domestic NNIE board chip and deployed it in the highway monitoring system of Suzhou City. (4) Outcomes: Published one first-author paper at CVPR24.

KEYVIA Intelligent Image Recognition for Substation Electrical Equipment 2022.9-2023.7

- **Project Leader** Addressed the technical challenge of **automatically identifying defects and abnormal states of substation electrical equipment during remote inspections**. (1) Data: Real-world inspection data from 120 types of indoor and outdoor locations, including data cleaning and boundary determination. (2) Models and Algorithms: Point location correction and positioning based on planar tracking. (3) Deployment: Verified across 160 high-speed rail substations under 10 railway bureaus nationwide, covering a total of 3,200 kilometers of railway. Achieved real-time performance (2.4Hz), multi-concurrent processing (second-level response), and an inspection accuracy of 96.3%. (4) Outcomes: Published one first-author paper at ICCV23.

GOVERNMENTAL RESEARCH GRANTS AS PROGRAM DIRECTOR

Adaptive Planar Target Perception via Multi-Level Amodal Reasoning Networks
NSFC - Young Students Fund of NSFC (PhD Student) - **300K RMB**

Digital Human Dialogue Video Generation via Multi-Dimensional Instruction Constraints
CIE & Tencent - PhD Student Incentive Program - **100K RMB**

HONORS AND AWARDS

• Young Elite Scientists Sponsorship Program by CAST (PhD Student Special Plan)	2025.01
• Nankai Top Ten Outstanding Students	2024.11
• Gold Award, Huawei Ascend AI Innovation Competition (Tianjin Division)	2024.11
• National Scholarship for Graduate Students (PhD Student)	2024.10
• CVPR'24 Outstanding Reviewer	2024.06
• National Scholarship for Graduate Students (PhD Student)	2023.10
• SK Artificial Intelligence Innovation Scholarship (First Prize)	2023.04
• University-Level Special Scholarship, Xidian University	2020.12

张知诚

☎ 185-6589-8403 · ✉ zhichengzhang@mail.nankai.edu.cn · 🌐 nku-zhichengzhang

教育背景

南开大学 博士生, 计算机科学与技术, 导师: 杨巨峰, 程明明 2021.9 - 现在 (预计毕业 2026.6)
西安电子科技大学 学士, 智能科学与技术 2017.9 - 2021.6

研究工作 (多模态大模型/视频图像生成/视频理解/平面重建与跟踪)

[Submit to NeurIPS25] VidEmo: Causal Affective-Cues Reasoning for Emotion-centric VideoLLM
Zhicheng Zhang, Weicheng Wang, Yongjie Zhu, Wenyu Qin, Pengfei Wan, Di ZHANG, Jufeng Yang 第一作者

[ICML25 SPOTLIGHT] MODA: MODular Duplex Attention for Multimodal Perception, Cognition, and Emotion Understanding
Zhicheng Zhang, Wuyou Xia, Chenxi Zhao, Zhou Yan, Xiaoqiang Liu, Yongjie Zhu, Wenyu Qin, Pengfei Wan, Di ZHANG, Jufeng Yang 第一作者

[CVPR24] ExtDM: Distribution Extrapolation Diffusion Model for Video Prediction
Zhicheng Zhang, Junyao Hu, Wentao Cheng, Danda Paudel, Jufeng Yang [PDF] [Code] [Project] 第一作者

[CVPR24] MART: Masked Affective RepresenTation Learning via Masked Temporal Distribution Distillation
Zhicheng Zhang, Pancheng Zhao, Eunil Park, Jufeng Yang [PDF] [Code] [Project] 第一作者

[CVPR23] Weakly Supervised Video Emotion Detection and Prediction via Cross-Modal Temporal Erasing Network
Zhicheng Zhang, Lijuan Wang, Jufeng Yang [PDF] [Code] [Video] 第一作者

[ICCV23] Multiple Planar Object Tracking
Zhicheng Zhang, Shengzhe Liu, Jufeng Yang [PDF] [Code] [Project] [MPOT-3K Dataset] [Video] [Demo] 第一作者

[TNNLS24] PlaneSeg: Building a Plug-in for Boosting Planar Region Segmentation
Zhicheng Zhang, Song Chen, Zichuan Wang, Jufeng Yang [PDF] [Code] 第一作者

[ACM MM22] Temporal Sentiment Localization: Listen and look in Untrimmed Videos
Zhicheng Zhang, Jufeng Yang [PDF] [Code] [TSL300 Dataset] [Video] 第一作者

[模式识别与人工智能 24] 属性知识引导的自适应视觉感知与结构理解研究进展
张知诚, 杨巨峰, 程明明, 林巍峤, 汤进, 李成龙, 刘成林 [PDF] 第一作者

[CVPR24] LAKE-RED: Camouflaged Images Generation by Latent Background Knowledge Retrieval-Augmented Diffusion
Pancheng Zhao, Peng Xu, Pengda Qin, Deng-ping Fan, Zhicheng Zhang, et al. [PDF] [Code] [Project] 学生二作

[中国图像图形学报 25] 情智兼备数字人与机器人研究进展
赵思成, 丰一帆, 张知诚, 孙斌, 张盛平, 高跃, 杨巨峰, 刘敏, 姚鸿勋, 王耀南 [PDF] 学生二作

实习经历

快手可灵 | Kling AI 2024.9-2025.5

- 主要负责人 解决数字人互动大模型中 Visual-Text 模态偏置问题。(1) Visual-Text 模态偏置问题：现有多模态大模型注意力存在对文本模态的偏置，容易忽略重要模态（视觉）的影响，丢弃 39% 重要的视觉线索。(2) 模型和算法：a. 基于注意力分解的双流自模态与跨模态注意力机制。b. 面向感知、认知与情感的多模态大模型。(3) 成果产出：中稿一作 ICML2025 论文，获评 SPOTLIGHT (2.6%)。
- 主要负责人 解决多模态人脸视频细粒度感知 Video2Text 技术问题。(1) 数据：a. 2w 小时人像视频数据，针对人像属性与情感动作，构建问答 QA 与高质量细粒度的 Caption 数据。b. 基于表情动作对齐的时序分段机制构造数据。(2) 模型和算法：a. 基于人像属性过程奖励模型的推理机制。b. 基于 Qwen2.5-VL 与 LLaVA-NeXT-OV 的 Caption 模型。(3) 成果产出：投稿一作 NeurIPS2025 论文，技术应用支持 kling2.0 语义响应业务线。

横向课题

- 华为 | HUAWEI 全天候视频语义理解技术合作项目2021.9-2022.8
- 凯发电气 | KEYVIA 变电站电气设备智能图像识别2022.9-2023.7

研究课题

- 基于多层级非模态推理网络的自适应平面目标感知
国家自然科学基金青年学生基础研究项目-博士研究生 (主持)
- 基于多维指令一致性约束的数字人对话视频生成
中国电子学会-腾讯博士生激励计划 (主持)
- 情智兼备的数字人与机器人
中国科协 2024 重大科学问题 (参与)
- 面向开放环境的自适应感知
科技创新 2030 “新一代人工智能” 重大项目 (参与)

荣誉与奖项

- 中国科协青年人才托举工程博士生专项计划2025.01
- 南开十杰2024.11
- 华为晟腾 AI 创新大赛天津赛区金奖2024.11
- 研究生国家奖学金（博士生）2024.10
- CVPR’24 杰出审稿人2024.06
- 研究生国家奖学金（博士生）2023.10
- SK 人工智能创新奖学金（一等奖）2023.04
- 西安电子科技大学校级特等奖学金2020.12