

# ZHICHENG ZHANG

☎ (+86) 185-6589-8403 · ✉ zhichengzhang@mail.nankai.edu.cn · 🌐 nku-zhichengzhang

## BIOGRAPHY

---

I am currently pursuing a Ph.D. degree at Nankai University, supervised by Prof. Jufeng Yang. My research interests include computer vision and deep learning, particularly focusing on **video understanding** and **video-language multimodal learning**. The main objective of my research is to enable machines to understand high-level instances as humans, including cognitive emotion and geometric primitive.

## EDUCATION

---

**Nankai University**, *Ph.D. student*, **Advisor:** Jufeng Yang 2021.9 - present  
**Xidian University**, *B.Eng.* 2017.9 - 2021.6

## RESEARCH

---

### ① *Temporal dynamics in analysing cognitive emotion*

**Weakly Supervised Video Emotion Detection and Prediction via Cross-Modal Temporal Erasing Network** [code]

2021.11-2022.3

- **Zhicheng Zhang**, Lijuan Wang, Jufeng Yang First author, **CVPR23**
- We introduce a weakly supervised network to exploit keyframes and the necessary context in a unified framework, which encourages the model to extract features from multiple discriminative parts. Our proposed method respectively obtains improvement of 3.2%, 2.9%, and 1.8% on three challenging benchmark datasets.
- To find the discriminative temporal part, we exploit intra- and inter-modal relationships to locate the key temporal segment, with which both holistic representations from the overall video and the local one from keyframes are fused.

**Temporal Sentiment Localization: Listen and look in Untrimmed Videos** [code] 2021.12-2022.4

- **Zhicheng Zhang**, Jufeng Yang First author, **ACM MM22**
- We propose a weakly-supervised framework to simultaneously locate and classify multiple emotional events, which may convey different categories in a video. When only using categorical labels in the keyframes ( $71.39\times$  less annotation burden than the full supervision), our method achieves an mAP of 28.72 and outperforms previous best weakly-supervised methods with a margin of 33.43%.
- Considering that multiple events are derived from multimodal stimuli, we propose a temporal sentiment distribution learning strategy and model the sentiment class conveyed in each segment according to the intensity.

### ② *Spatial dynamics in recognizing geometric planes*

**PlaneSeg: Building a Plug-in for Boosting Planar Region Segmentation** [code] 2021.3-2022.3

- **Zhicheng Zhang**, Song Chen, Zichuan Wang, Jufeng Yang First author, **TNNLS23**
- We design a plug-in framework that provides edge-aware features to address the problem of ambiguous boundaries. When integrated into existing methods, our framework brings an average improvement of 7.42%, 3.70%, 10.86%, and 3.67% over PlaneRecNet on plane segmentation, plane detection, plane reconstruction, and depth prediction, respectively.

**Multiple Planar Object Tracking** [code] 2022.5-2023.3

- **Zhicheng Zhang**, Jufeng Yang First author, **ICCV23**
- We propose a tracking-by-reasoning framework to track both location and pose of the planar object. Our framework consolidates occlusion reasoning as well as appearance perception, and outperforms SOTAs with a margin of 1.84% and 16.9% for single and multiple planar object tracking.
- To address the complex occlusion raised by relative movement between planar objects, we propose an occlusion area localization strategy based on the historical motion of planar objects to indicate the area explicitly.

## HONORS AND AWARDS

---

- SK AI Innovation Scholarship (the First Pride) 2023.4
- Nankai Gongneng Scholarship 2021.9
- Xidian Special Scholarship 2020.12