Statement of Purpose

Name: Yuanhao Cai
Telephone: +86-15652599567
Homepage Github Email

To be a creative scientist for the world

I want to pursue a Ph.D. because I like doing research and enjoy solving valuable problems. My career aspiration is to become a professor and establish my lab.

My research interests include computer vision and deep learning, mainly focusing on image and video restoration like snapshot compressive imaging [1, 2, 3, 6], image denoising [4, 9, 10], spectral reconstruction [1, 12], medical image enhancement [8], video deblurring [7, 11], etc. I believe these topics are valuable because the first step human obtain information from vision is to perceive. However, some low-end imaging equipments or poor conditions may constrain our observation of the desired scenes.

Thus, my goal is to recover these underlying scenes and enhance their quality, which can help people obtain more comprehensive and accurate information.

Totally, I have three research experiences. When I was a junior (2018 - 2019), I got a chance to participate in the RoboCup world final as a member of the Tsinghua University team. I was in charge of robot vision. The main task was to detect eight kinds of objects (e.g., football, goal, etc.) in the video captured by the robot camera. My job included collecting data, labeling images, and training and deploying models on hardware. Eventually, our team won two runner-up awards. This was my first project of computer vision. I was deeply attracted by its power and practicality.

From 2019 to 2020, I focused on studying human pose estimation. I worked as an intern in Megvii, supervised by professor Jian Sun, who won the CVPR best paper award twice. At that time, existing methods mainly focused on learning global semantic features by aggregating inter-level representations, resulting in the lack of delicate local spatial information and leading to imprecise keypoint detection. To tackle this issue, I proposed a novel multi-scale architecture simultaneously capturing local and global features by intra- and inter-level fusion. Besides, I also customized a pose refinement mechanism to adaptively adjust the output representations. Eventually, I presented my method in the top conference ECCV 2020 as Spotlight [5]. Besides, I also used my methods to win the COCO Keypoint Detection Challenge twice in ICCV 2019 and ECCV 2020 as the first and co-first author. I learned a lot from this experience. Not only did I improve my research skills, but also the experience shaped my toughness and confidence in academic research.

In my Master's study (since 2020), I focus on studying image and video restoration problems. I study the degradation type, pattern, and distribution of real imaging scenes. I proposed the first Transformer-based methods for snapshot compressive imaging [1], spectral reconstruction [12], fundus image enhancement [8], and video deblurring [7].

The work I am particularly proud of is my CVPR 2022 paper MST [1]. In this work, I resolved the spectral compressive imaging reconstruction problem with deep insight. Specifically, I combined the spatially sparse while spectrally similar nature of hyperspectral images with the computation paradigm of Transformer. Different from the original Transformer that treats each spatial pixel vector as a token and suffers from enormous computational complexity, my MST treats a spectral feature map as a token and computes the self-attention along the spectral dimension. This novel scheme can reduce tons of computational costs and better fit to the property of hyperspectral images. My method significantly outperformed previous methods by over 2.5 dB while requiring only less than 5% of computational complexity and

model parameters.

In addition, my further work MST++ [12] based on MST won the first place of NTIRE 2022 Spectral Reconstruction Challenge. The NTIRE series of competitions is the most authoritative and challenging competition in low-level vision. During this competition, I demonstrated excellent leadership, cooperation awareness, and strong persistence. I firstly designed the whole algorithm framework and technology roadmap, and then assigned tasks to each team member according to their strengths and skills. I organized meetings per week to discuss ideas and checked the progress of each team member. Everyone cooperated efficiently. When It was only a week before the deadline, we encountered a bug that limited the performance but we did not know what the problem was. There was a big gap between our team and the first-ranked team. Under immense pressure, I worked for almost 16 hours per day to debug. Finally, I found that the pre-normalization of data was wrong. The data should be normalized to the range of $0 \sim 1$. However, the data was mapped into the range of $-1 \sim 1$. After fixing this bug, we surpassed other teams and won the first place.

Furthermore, my open-source work about this paper is solid. I released the codes of MST and MST++ on Github. I extended the two repositories to two toolboxes for spectral compressive imaging and spectral reconstruction from RGB images. Each of them includes over 10 deep learning based methods for comparison. The two repositories received over 600 stars in eight months. I believe solid open-source work can gradually and effectively increase the impact of my papers.

Although I have many papers published by top conferences, I realize that there are still many critical problems need to be solved in the field of image restoration. Thus, I decide to do a Ph.D. after finishing my Master's study. I believe my high motivation level, proficient skills, solid research background, and academic collaboration connections would promise my success in the Ph.D. program.

My research plan can be summarized into four points. Firstly, I would like to conduct further research based on previous works. I plan to study the efficiency problem of image and video restoration. Current methods usually pursue better performance by employing enormous computational and memory costs. My goal is to design deep learning algorithms with higher performance, lower computation, and fewer parameters for image or video quality enhancement. In my opinion, a solution is combine the computation paradigm of Transformer and specific degradation types to develop efficient algorithms. In particular, I plan to expand my work MST [1], CST [3], and DAUHST [2] into top journal papers like TPAMI. Secondly, I would like to investigate other under-explored topics of image restoration like low-light image enhancement and blind image super-resolution. There are still no effective solutions for these severe ill-posed tasks because of their tough degradation patterns. I plan to leverage the generative capacity of diffusion model to simulate the pattern and distribution of real-camera imaging degradation. Thirdly, I would like to enlarge my research scope from 2D vision to 3D vision, i.e., to enhance the quality of 3D point clouds such as point cloud up-sampling, completion, and denoising. 3D vision has wide applications such as autopilot, game and animation rendering, etc. Point cloud is one of the most important 3D vision representations. However, point clouds obtained by LIDAR are sparse, noisy, and incomplete. This hinders the perception of 3D scenes. Thus, studying how to enhance the quality of 3D point clouds is valuable. Finally, I will study how to use low-level vision techniques to benefit the high-level vision tasks like object detection, semantic segmentation, etc. For example, the visibility at night can be very low, which poses difficulties to vehicle and pedestrian detection. But if the low-light images can be enlightened first, the tasks would be much easier.

References

[1] **Yuanhao Cai** *, Jing Lin *, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided Spectral-wise Transformer for Efficient Hyperspectral Image Reconstruction. ***IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)***, 2022

[2] **Yuanhao Cai** *, Jing Lin *, Haoqian Wang, Xin Yuan, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Degradation-Aware Unfolding Half-Shuffle Transformer for Spectral Compressive Imaging. ***Advances in Neural Information Processing Systems (NeurIPS)***, 2022

[3] **Yuanhao Cai** *, Jing Lin *, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. ***Coarse-to-Fine Sparse Transformer for Hyperspectral Image Reconstruction. European Conference on Computer Vision (ECCV)***, 2022

[4] **Yuanhao Cai** , Xiaowan Hu, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, and Donglai Wei. Learning to Generate Realistic Noisy Images via Pixel-level Noise-aware Adversarial Training. ***Advances in Neural Information Processing Systems (NeurIPS)***, 2021

[5] **Yuanhao Cai** *, Zhicheng Wang *, Zhengxion Luo, Binyi Yin, Ang'ang Du, Haoqian Wang, Xiangyu Zhang, Xinyu Zhou, ErJin Zhou, and Jian Sun. ***Learning Delicate Local Representations for Multi-Person Pose Estimation. European Conference on Computer Vision (ECCV), Spotlight***, 2020

[6] Xiaowan Hu *, **Yuanhao Cai** *, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. HDNet: High-resolution Dual-domain Learning for Spectral Compressive Imaging. ***IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)***, 2022

[7] Jing Lin *, **Yuanhao Cai** *, Xiaowan Hu, Haoqian Wang, Youliang Yan, Xueyi Zou, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Flow-Guided Sparse Transformer for Video Deblurring. ***International Conference on Machine Learning (ICML)***, 2022

[8] Zhuo Deng *, **Yuanhao Cai** *, Lu Chen, Zheng Gong, Qiqi Bao, Xue Yao, Dong Fang, Shaochong Zhang, and Lan Ma. RFormer: Transformer-based Generative Adversarial Network for Real Fundus Image Restoration on A New Clinical Benchmark. ***IEEE Journal of Biomedical and Health Informatics (J-BHI)***, 2022

[9] Xiaowan Hu, Ruijun Ma, Zhihong Liu, **Yuanhao Cai**, Xiaole Zhao, Yulun Zhang, and Haoqian Wang. Pseudo 3D Auto-Correlation Network for Real Image Denoising. ***IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)***, 2021

[10] Xiaowan Hu, **Yuanhao Cai**, Zhihong Liu, Haoqian Wang, and Yulun Zhang. Multi-Scale Selective Feedback Network with Dual Loss for Real Image Denoising. ***International Joint Conference on Artificial Intelligence (IJCAI), Oral***, 2021

[11] Jing Lin *, Xiaowan Hu *, **Yuanhao Cai**, Haoqian Wang, Youliang Yan, Xueyi Zou, Yulun Zhang, and Luc Van Gool. Unsupervised Flow-Aligned Sequence-to-Sequence Learning for Video Restoration. ***International Conference on Machine Learning (ICML)***, 2022

[12] **Yuanhao Cai** *, Jing Lin *, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte,

and Luc Van Gool. MST++: Multi-stage Spectral-wise Transformer for Efficient Spectral Reconstruction. ***IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)***, 2022