

Zhou Daquan

RESEARCH SCIENTIST AT BYTEDANCE USA RESEARCH LAB

☎ (065) 93387136 | ✉ zhoudaquan21@gmail.com | in Zhou Daquan | 8 Zhou Daquan



Google scholar Citation: 5020
(Co-) First Author Paper: 12
Co-author papers: 25
H-index: 20

Research Interest

As an ardent researcher, I am keen on consuming the minimum energy & memory consumption for training & using the most powerful artificial intelligence algorithms. I am a staunch advocate for green AI and the advancement of general artificial intelligence. My prior research has extensively focused on the efficiency and generalization of deep neural networks, as well as hardware execution efficiency. This expertise positions me uniquely to further delve into research that converts powerful AI algorithms from useful into both powerful and usable in the realm of edge devices. I have been focused on the video-generative model recently and I am also extremely interested in exploring utilizing video-generative pre-trained models as world models and using them for interaction with the environment.

Research Experience

Bytedance

RESEARCH SCIENTIST, AI LAB US

San Jose, USA

July, 2022 - Now

- Focus on the research works of architecture design with efficiency consideration for the image/video generation and multi-modality large language model (LMM) tasks.

Nvidia

INTERNSHIP, NVIDIA RESEARCH

Mountain View, USA

June 2021 - March 2022

- Explore design guidelines on the neural network architectures that are robust to nature image corruptions with zero shot training framework.
- Focus on the design of efficient neural architectures with strong zero-shot robustness capabilities.
- Proposed algorithms are deployed in the autonomous car and are the core algorithm for the Champion solutions for the Robust Vision Challenge 2022 for five segmentation tracks.

NUS - IDS interdisciplinary program

PHD STUDENT, NUS, LEARNING AND VISION LAB

Singapore

Aug. 2018 - 2022 May

- Focus on research related to deep learning model compression and acceleration, AutoML, and hardware-algorithm co-design framework.
- Explore efficient deep neural network architectures and deployment on edge devices.
- Advised by Dr. Jiashi Feng

DSO - ST Electronics Joint Lab

SYSTEM DESIGN ENGINEER

Singapore

June 2016 - August 2018

- Focus on the design and develop a high-speed real-time operation system with FPGA - Micro-processor co-processing.
- Launch the first commercial satellite in Singapore (Singapore President Science and Technology Award (Group Award))

Education Background and Awards

National University of Singapore

PHD STUDENT OF INSTITUTE OF DATA SCIENCE

Singapore

June 2018 to June 2022

- Google Scholar Citations: 5020
- Top tier conference: 25 papers (12 as first authors) accepted by ICML, ECCV, NeurIPS, ICLR, ICCV, and CVPR
- Top tier journal: 4 papers accepted by T-PAMI, TIP and TNNLS
- 2020 SINGAPORE DATA SCIENCE CONSORTIUM (SDSC) DISSERTATION RESEARCH FELLOWSHIP AWARD

National University of Singapore

BACHELOR OF ENGINEERING IN ELECTRICAL AND COMPUTER ENGINEERING

Singapore

June 2012 to June 2016

- Graduate with **Highest distinction**, first class honors degree
- **Dean's List ×4**, Top Academic Performance Awards
- **Attained Singapore President Science and Technology Award**, Group Award, STE Satellite System — *Fall 2017*
- **Awarded Infineon Technology Asia Pacific Prize**, — *Top-1 student in the IC design specialization track*
- **Awarded best intern prize**, SEAGATE TECHNOLOGY — *June 2014*
- **Attained First Runner-Up for AI design of 5-on-5 Robot soccer simulation competition**, — *Spring 2013*
- **Awarded Scholar of China and Singapore Ministry of Education scholarship**, — *Fall 2012*

Research Activities

CVPR2022 Workshop: Transformers for Vision

Vancouver, Canada

CORE ORGANIZING COMMUNITIES

June, 2022

- Organizing the first workshop on the new backbone Transformer
- Jointly organize the workshop with researchers from Google Brain, Meta AI, Nvidia Research, MSR, UT Austin and CalTech.

Invited Talk at Toronto University AI Lab

Vancouver, Canada

KEY SPEAKER

Dec, 2023

- A talk on video generation and customerization
- Jointly invited by Nvidia Research Team

Collaboration with Academic Institutes

San Jose, USA

MENTORING STUDENTS FROM ACADEMIC INSTITUTES FOR RESEARCH PROJECTS

March 2023 to Now

- Collaboration with Peking Universities, mentoring 4 students on research projects regarding deep neural network efficiency, large language model (LLM) and IP protection algorithm.
- Collaboration with Nankai Universities, mentoring 2 students on research projects regarding AIGC.
- Collaboration with UC Berkeley, mentoring 2 students on research projects regarding AIGC and multi-agents.

Invited Talk at Workshop in IJCAI

Macao, China

AS A DISTINGUISHED SPEAKER ON THE WORKSHOP WITH TOPICS OF MODEL GENERALIZATION AND EFFICIENCY.

August, 2023

Invited talk from Hua Wei

Hang Kong, China

INVITED FOR A TALK ON DIFFUSION ALGORITHMS FOR IMAGE AND VIDEO GENERATION

December, 2022

Guest Lecturer

Beijing, China

INVITED AS A LECTURER FOR A MODULE AT THE UNIVERSITY OF CHINESE ACADEMY AND SCIENCE.

April, 2023

Peer Review Experience

REVIEWER FOR NEURIPS, ICML, CVPR, ICCV, ICLR, T-PAMI, TNNLS AND NEROComputing

Invited Talk at TechBeat

Beijing, China

INVITED TALK ON TRANSFORMER MODEL ARCHITECTURE DESIGN AND GENERALIZATION CAPABILITY, AWARDED AS THE MOST

POPULAR TALK PRESENTER

Jan 2021 & March 2022

Selected Publications

Publications Summary

Efficiency, Generalization & Neural operator design

- Highlights: **3 oral presentations + 3 spotlight presentations + 1 Best paper nomination + 1 CVPR Most Influcial Paper (Ranked 5)**
- New area exploration:
 - Propose the first text to video generation framework (MagicVideo) in latent space that:
 - * Reduces the computation cost by 64×
 - * Followed by prestigious researchers from:
 - Universities: Stanford, Princeton, Toronto, etc.
 - Companies: Nvidia Research, Google Brain, Meta AI, Microsoft Research, etc.
- Impacts in realm:
 - Proposed architecture (FAN, ICML2022) is deployed in the autonomous driving vision system.
 - Proposed efficient diffusion algorithm (SlimDiff, CVPR 2023) is deployed in the products from Bytedance.
 - Proposed architecture (FAN, ICML 2022) is used as the backbone for the champion solution in the Robust Vision Challenge (RVC) in conjunction with ECCV 2022.

Generative Model Design

MagicVideo: Efficient Video Generation With Latent Diffusion Models

- Daquan Zhou, Weimin Wang, Hanshu Yan, Weiwei Lv, Yizhe Zhu, Jiashi Feng
- Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI) (Under review)

MagicVideoV2: Long Video Generation and Editing via Sliding Window Temporal-attention

- Daquan Zhou, Jiashi Feng
- Preprint

Expanding small-scale datasets with guided imagination

- Yifan Zhang, Daquan Zhou*, Bryan Hooi, Kai Wang, Jiashi Feng
- Correspondence author, project lead
- Accepted by Conference on Neural Information Processing Systems (NeurIPS) 2023

Diffusion probabilistic model made slim

- Xingyi Yang, Daquan Zhou, Jiashi Feng, Xinchao Wang
- Accepted by Computer Vision and Pattern Recognition Conference (CVPR) 2022

Efficiency

Dataset Quantization

- Daquan Zhou, Kai Wang, Jianyang Gu, Xiangyu Peng, Dongze Lian, Yifan Zhang, Yang You, Jiashi Feng
- Accepted by ICCV, or the International Conference on Computer Vision (ICCV) 2023.
- A new dataset compression pipeline with 5× to 10× training speed up.

Scaling & Shifting Your Features: A New Baseline for Efficient Model Tuning

- Dongze Lian*, Daquan Zhou*, Jiashi Feng, Xinchao Wang
- Accepted by Conference on Neural Information Processing Systems (NeurIPS) 2022, **spotlight presentation, Euqual First Author**

EPIM: Efficient Processing-In-Memory Accelerators based on Epitome

- Chenyu Wang*, Zhen Dong*, Daquan Zhou*, Zhenhua Zhu, Yu Wang, Jiashi Feng, Kurt Keutzer
- Accepted by Design Automation Conference (DAC) 2023, WIP. **Equal first author.**
- Collaboration with scholars from UC Berkeley for hardware-algorithm co-design project.

Rethinking Bottleneck Structure for Efficient Mobile Network Design

- Daquan Zhou, Qibin Hou, Yunpeng Chen, Jiashi Feng, Shuicheng Yan
- Accepted by European Conference on Computer Vision (ECCV) 2020, spotlight presentation

Neural Epitome Search for DNN Model Compression

- Daquan Zhou, Xiaojie Jin, Qibin Hou, Kaixin Wang, Jianchao Yang, Jiashi Feng
- Accepted by International Conference on Learning Representations (ICLR) 2020

Coordinate Attention for Efficient Mobile Network Design

- Qibin Hou, Daquan Zhou, Jiashi Feng
- Accepted by Computer Vision and Pattern Recognition Conference (CVPR) 2020
- **Ranked 5 among Most Influential Paper in CVPR**

Deep Model Reassembly

- Xingyi Yang, Daquan Zhou, Songhua Liu, Jingwen Ye, Xinchao Wang
- Accepted by Conference on Neural Information Processing Systems (NeurIPS) 2022, **best paper nomination**

Convbert: Improving bert with span-based dynamic convolution

- Zihang Jiang, Weihao Yu, Daquan Zhou, Yunpeng Chen, Jiashi Feng, Shuicheng Yan
- Accepted by Conference on Neural Information Processing Systems (NeurIPS) 2020

Model Generaliation

DeepViT: Towards Deeper Vision Transformer

- Daquan Zhou, Bingyi Kang, Xiaojie Jin, Linjie Yang, Xiao Chen Lian, Qibin Hou, Jiashi Feng
- Accepted as a journal paper in IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI) 2022

Understanding The Robustness in Vision Transformers

- **Daquan Zhou**, Zhiding Yu, Enze Xie, Chaowei Xiao, Anima Anandkumar, Jiashi Feng, Jose M. Alvarez
- Accepted by International Conference on Machine Learning (ICML) 2022 (**short oral presentation**)
- Champion solutions for Robust Vision Challenge 2022

AutoSpace: Neural Architecture Search with Less Human Interference

- **Daquan Zhou**, Xiaojie Jin, Xiaochen Lian, Linjie Yang, Yujing Xue, Qibin Hou, Jiashi Feng
- Accepted by ICCV, or the International Conference on Computer Vision (ICCV) 2021

Shunted Self-Attention via Multi-Scale Token Aggregation

- Sucheng Ren*, **Daquan Zhou***, Shengfeng He, Jiashi Feng, Xinchao Wang
- Accepted by IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) 2022, **Oral Presentation, Equal First Author**

Velocity-to-velocity human motion forecasting

- Hongsong Wang, Liang Wang, Jiashi Feng, **Daquan Zhou**
- Accepted by 2022 IEEE Transactions on Image Processing (TIP) 2022

PANet: Few-Shot Image Semantic Segmentation with Prototype Alignment

- Kaixin Wang, Jun Hao Liew, Yingtian Zou, **Daquan Zhou**, Jiashi Feng
- Accepted by IEEE/CVF International Conference on Computer Vision (ICCV) 2019