JUNTANG ZHUANG

Email: zhuangjt12@gmail.com ⋄ Website: https://juntang-zhuang.github.io ⋄ Phone: +1 475-224-8282 ⋄ [Google Scholar]

ABOUT ME

I am a Research Sceintist at OpenAI. I obtained my Ph.D. degree in Biomedical Engineering from Yale University in 2021, advised by Prof. James S. Duncan. My research interest lies in machine learning, optimization, large language models and numerical methods.

EDUCATION

Yale University
Ph.D. in Biomedical Engineering (Advisor: James S. Duncan) — Sep 2016 - April 2022

Yale University
M.A. in Statistics, M.Phil in Biomedical Engineering — Sep 2017 - May 2018

Tsinghua University
B.E. in Engineering Physics — Sep 2012 - May 2016

WORK EXPERIENCE

OpenAI — San Francisco, C.A. April 2022 - now
Student researcher at Google — Los Angeles, C.A. June-Oct 2021

PROJECTS

GPT-next: contributor to next-generation GPT

GPT4-Turbo: inventor of long-context algorithm supporting 256k window

OpenAI Embedding v3: core contributor

DALL·E 2.5 and DALL·E 3: primary contributor

GPT-4: co-author

OPEN-SOURCE PROJECTS

AdaBelief-optimizer (*1k* stars on github, added to official repositories such as Deepmind optax, Tensorflow-Addons and Google Flax); ShelfNet (246 stars); LadderNet (102 stars); TorchDiffEqPack

AWARDS & SCHOLARSHIPS

- Henry Prentiss Becton Graduate Prize (1 out of Yale School of Engineering & Applied Science) 2022
- Best paper award, Machine Learning in Medical Imaging (MLMI) 2019
- Top-1 winner for CNI Transfer Learning Challenge, MICCAI 2019
- Graduate fellowship, Yale University 2016
- Award for excellent learning performance, Tsinghua University 2015
- Meritorious award for Mathematical Contest in Modeling (top 10% teams worldwide) 2015
- National encouragement award (for excellent learning performance), Tsinghua University 2014
- Sparks Program (Undergraduate High-tech Club) membership, Tsinghua University 2014

PUBLICATIONS

1. **J. Zhuang**, Boqing Gong, et al., Surrogate gap minimization improves sharpness-aware training *International Conference on Learning Representaitions* (ICLR 2022)[project page]
2. **J. Zhuang**, Yifan Ding, et al., Momentum centering and asynchronous update for adaptive gradient methods *Conference on Neural Information Processing Systems* (NeurIPS 2021)[project page]
3. **J. Zhuang**, N. Dvornek, et al. MALI: a memory efficient and reverse accurate integrator for Neural ODEs, *International Conference on Learning Representations* (ICLR 2021)[project page]
4. **J. Zhuang**, N. Dvornek, et al. Multiple-shooting adjoint method for whole-brain dynamic causal modeling , *Information Processing in Medical Imaging* (IPMI 2021, oral presentation) [project page]
5. **J. Zhuang**, T. Tang, et al. AdaBelief Optimizer: adapting stepsizes by the belief in observed gradients, *Conference on Neural Information Processing Systems* (NeurIPS 2020, Spotlight) [project page]

6. **J. Zhuang**, N. C. Dvornek, et al. Adaptive Checkpoint Adjoint Method for Gradient Estimation in Neural ODE, *International Conference on Machine Learning* (ICML 2020) [project page]
7. **J. Zhuang**, J. Yang, et al., ShelfNet for fast semantic segmentation, *Workshop on Computer Vision for Road Scene Understanding and Autonomous Driving* (CVRSUAD 2019)
8. **J. Zhuang**, N. C. Dvornek, et al., Decision Explanation and Feature Importance for Invertible Networks, *Workshop on Interpretating and Explaining Visual Artificial Intelligence Models* (XAIC 2019) [*oral*]
9. **J. Zhuang**, N. C. Dvornek, et al., Invertible Network for Classification and Biomarker Selection for ASD , *International Conference on Medical Image Computing & Computer Assisted Intervention* (MICCAI 2019)
10. **J. Zhuang**, N. C. Dvornek, et al., Prediction of Pivotal response treatment outcome with task fMRI using random forest and variable selection, *International Symposium on Biomedical Imaging* (ISBI 2018)
11. **J. Zhuang**, N. C. Dvornek, et al., Prediction of severity and treatment outcome for ASD from fMRI, *International Workshop on PRedictive Intelligence In MEdicine* (PRIME 2018)
12. N. C. Dvornek, X. Li, **J. Zhuang**, et al., Jointly discriminative and generative recurrent neural networks for learning from fMRI (MLMI 2019)
13. **J. Zhuang** LadderNet: Multi-path networks based on U-Net for medical image segmentation, ArXiv 2018

RESEARCH EXPERIENCE

1. Optimization for deep learning

Surrogate gap minimization improves sharpness-aware training Jun - Oct 2021
Mentor: Ting Liu, Boqing Gong et al. Google Research

- Proposed a generic method to improve the generalization of neural networks. Provided extensive theoretical analysis on both the convergence and provably better generalization performance of the new method.
- Empirically validated that the proposed method consistently improves the test performance of neural networks. Specifically, for the ImageNet top-1 accuracy the proposed method achieved **+11.3%** improvement over AdamW on Vision Transformer, and **+12%** improvement on MLP-Mixer.
- Paper accepted to ICLR 2022.[project page]

Momentum centering and asynchronous update for adaptive gradient methods Jan - May 2021
Mentor: James S. Duncan, Sekhar Tatikonda Yale University

- Proposed ACprop, which is an adaptive optimizer combining momentum centering and asynchronous update. Theoretically, ACProp has the optimal convergence rate and weak convergence conditions.
- Validated ACProp in extensive empirical studies: ACProp outperforms both SGD and other adaptive optimizers in image classification with CNN, GAN models, reinforcement learning and transformers.
- Paper accepted to NeurIPS 2021. [project page]

AdaBelief optimizer: a fast, accurate and stable optimizer for deep learning Jan - June 2020
Mentor: James S. Duncan, Sekhar Tatikonda Yale University

- Developed an optimizer for deep learning models. To our knowledge, it's the first to achieve three goals simultaneously: *fast training speed, good generalization performance, and stability of training.*
- Performed extensive validation in computer vision tasks (+3% on ImageNet top-1 accuracy), language modeling and generative adversarial networks (GAN) (e.g. -0.5 in FID score) and reinforcement learning.
- Paper accepted as **Spotlight Presentation** by NeurIPS 2020. [project page]

2. Solvers for continuous-time neural networks

MALI: a memory efficient and reverse accurate integrator for Neural ODEs Sep - Nov 2020
Mentor: James S. Duncan, Sekhar Tatikonda Yale University

- Proposed MALI, a new solver for Neural ODEs with numerical accuracy at a constant memory cost.
- MALI is suitable for large-scale Neural ODEs, such as FFJORD, and large ODE-CNNs on ImageNet. MALI achieves new state-of-the-art (3.71 BPD on ImageNet64) for image generation with continuous models.
- Paper accepted by International Conference on Learning Representations (ICLR 2021) [project page]

Adaptive checkpoint adjoint method for gradient estimation in neural ODE Jun - Dec 2019
Mentor: James S. Duncan, Sekhar Tatikonda Yale University

- Proposed and implemented a family of adaptive ODE solvers for accurate gradient estimation. Achieved both accuracy and computation efficiency. To our knowledge, our method is the first to enable neural-ODE to achieve comparable results to state-of-the-art discrete-layer models on benchmark classification tasks.
- Paper accepted by International Conference on Machine Learning (ICML 2020). [project page]

3. Prior-informed machine learning and biomedical applications

Evolutionary causal modeling of brain states from task-fMRI data Mar - Sep 2020
Mentor: James S. Duncan Yale University

- Modeled the effective connectome of the brain, which is the directional influence between different regions of the brain. Developed a differential equation model to simulate the dynamical evolution of brain states.
- Developed a method to efficiently learn the dynamical model from high-dimensional fMRI data.
- Paper accepted as **Oral Presentation** by 27th International Conference on Information Processing in Medical Imaging (IPMI 2021).

Invertible networks for model decision interpretation Jan - June 2019
Mentor: James S. Duncan, Nicha C. Dvornek Yale University

- Proposed a two-stage model for classification tasks, an invertible transform from input domain to feature domain, and a linear classifier in the feature domain. With invertible networks, we explicitly determine the decision boundary in the input domain, and calculate the projection of a point onto the decision boundary. The difference between a point and its projection onto the decision boundary can be viewed as the explanation for model decision.
- Applied the proposed method on fMRI data, selected biomarkers for ASD, and validated biomarkers in prediction of phenotype scores.
- Papers accepted by ICCV Workshop (ICCV 2019, XAIC) and MICCAI 2019.

ShelfNet for real-time semantic segmentation Aug - Nov 2018
Mentor: Nicha C. Dvornek Yale University

- Proposed ShelfNet, a multi-path network with a shelf-like structure for real-time semantic segmentation. Achieved both faster running speed and higher mIoU (e.g. 84.2 mIoU at 42 FPS on PASCAL VOC) than state-of-the-art real-time semantic segmentation models such as BiSeNet.
- Paper accepted by ICCV Workshop on Computer Vision for Road Scene Understanding and Autonomous Driving (ICCV 2019, CVRSUAD).

Treatment outcome prediction and biomarker selection for autism from fMRI 2017-2018

- Designed a two-level feature selection approach for the whole-brain image to deal with the high dimensionality: first select predictive ROIs, then select predictive voxels within surviving ROIs. Developed interpretable and accurate predictive models.
- Tested on both task fMRI and resting-state fMRI datasets and achieved state-of-the art accuracy on different clinical scores. Selected biomarkers for autism and validated the results with Neurosynth decoder.
- Paper accepted in PRIME workshop, MICCAI 2018 and ISBI 2018.

PROFESSIONAL ACTIVITY

Served as reviewer for MIDL 2019, ICML 2021, NeurIPS 2021, ICLR 2022, EMBC 2021 and MEDIA.

INVITED TALKS

- Computational Neuroscience Laboratory at Stanford University 2020
- SyncedTech (name in Chinese pinyin: JiQi ZhiXin) [video] 2020

SKILLS

C/C++, R, Python, MATLAB, PyTorch, Keras, Jax, Tensorflow