Jiatao Gu

Machine Learning Researcher, Apple | Ph.D.

Homepage: https://jiataogu.me/
Mobile: +1 (650) 885-1592
Work Email: jiatao@apple.com | Personal Email: thomagram@gmail.com
Google Scholar | LinkedIn | GitHub | Twitter

RESEARCH INTERESTS

Generative AI, Natural Language Processing, Multi-model Representation Learning, Embodied AI

EDUCATION

Ph.D in Electrical and Electronic Engineering   Aug 2014 – Sep 2018
The University of Hong Kong   Hong Kong, China
Supervisor: *Prof. Victor O.K. Li*
Dissertation: *Efficient Neural Machine Translation*

Bachelor in Electronic Engineering   Sep 2010 – Jun 2014
Tsinghua University   Beijing, China
GPA: 3.76, Rank: 18/243

RESEARCH EXPERIENCE

Staff Research Scientist   Jun 2022 – Present
Machine Learning Research (MLR), Apple Inc.   New York City, NY, US

Senior Research Scientist   Aug 2018 – Jun 2022
Facebook AI Research (FAIR), Meta Platforms Inc.   New York City, NY, US

Research Intern   Sep 2017 – Dec 2017
Salesforce Research, Salesforce Inc.   Palo Alto, CA, US
Mentor: *Dr. Richard Socher*

Research Intern   Apr 2017 – Aug 2017
Microsoft Research, Microsoft Corporation   Redmond, WA, US
Mentor: *Dr. Hany Hassan*

Visiting Scholar   Jun 2016 – Jan 2017
Courant Institute of Mathematical Science, New York University   New York City, NY, US
Adviser: *Prof. Kyunghyun Cho*

Exchange Student   Oct 2012 – Mar 2013
Graduate School of Information Science and Technology, The University of Tokyo   Tokyo, Japan
Adviser: *Prof. Hitoshi Iba, Danushka Bollegala*

PROFESSIONAL ACTIVITIES

Professional Service

> *Action Editor:* ACL Rolling Review (ARR), since 2022 - now
>
> *Area Chair:* ICLR 2025, 2024, ACL 2021, EMNLP 2020, 2021
>
> *Reviewer:* NeurIPS, ICML, ICLR, ACL, EMNLP, NAACL, IJCNLP, CVPR, ECCV, SIGGRAPH, SIGGRAPH Asia, AAAI, IJCAI

Courses

> *Non-autoregressive Sequence Generation*
> Cutting-edge Tutorial, ACL 2022   May 22, 2022

Organizer

> *Simultaneous Speech Translation Task*
> IWSLT 2020

Talks

Towards Flexible, Scalable and Knowledgable Generative Intelligence

New York University, guest lecture	Apr 16, 2024
University of Pennsylvania, guest lecture	Oct 16, 2024

BOOT: data-free distillation of denoising diffusion models with bootstrapping

ICML 2023 Workshop on Structured Probabilistic Inference & Generative Modeling   Jul 28, 2023

Towards Efficient Diffusion Models for 3D Generation

BIRS Workshop, 3D Generative Models, virtual talk   Jul 14, 2023

Non-autoregressive Sequence Generation

ACL 2022 Tutorial (https://nar-tutorial.github.io/acl2022/)   May 22, 2022

StyleNeRF:A style-based 3d-aware generator for high-resolution image synthesis

VALSE Webinar, virtual talk   Dec 29, 2021

Multilingual Denoising Pre-training for Neural Machine Translation

Google Translate, virtual talk	Jul 17, 2020
ACL2020 WNGT Workshop, virtual talk	Jul 10, 2020

Understanding Knowledge Distillation in Neural Sequence Generation

Tsinghua University, Beijing, China	Jan 03, 2020
Microsoft Translate, Redmond, WA, USA	Dec 04, 2019

Non-autoregressive Neural Machine Translation

New York University, New York, NY, USA	Dec 19, 2017
Google Brain, Mountain View, CA, USA	Nov 20, 2017

Learning to Share: towards Zero-Resource Neural Machine Translation without Pivots

Microsoft Research, Redmond, WA, USA   Jul 27, 2017

Trainable Decoding for Neural Machine Translation

NYU Shanghai, Shanghai, China   Apr 14, 2017

Learning to Translate in Real-time with Neural Machine Translation

Carnegie Mellon University, Pittsburgh, PA, USA	Oct 27, 2016
New York University, New York, NY, USA	Oct 19, 2016

Teaching Experience

Teaching Assistant

Teaching Assistant, ELEC 2501/3541 S/W Engineering and O.S., HKU	Spring, 2015
Teaching Assistant, ELEC 3226/3442 Embedded System, HKU	Spring, 2016
Teaching Assistant, ELEC 3641 Human computer interaction, HKU	Spring, 2018

Media Coverage

Reports mentioned **Matryoshka Diffusion**

http://bit.ly/3SwDVuL (Marktechpost)

Reports mentioned **Data2vec**

https://bit.ly/3Gr3Vhq (MIT Technology Review)

Reports mentioned **StyleNeRF: A 3D-Aware Generator for High-Resolution Image Synthesis**

https://bit.ly/35OeT49 (SyncedReview)

Reports mentioned **Multilingual Fine-tuning of Extensible Multilingual Pretraining**

https://bit.ly/3gpt01I (Towards data science)

Reports mentioned **Multilingual Denoising Pre-training for Neural Machine Translation**

https://bit.ly/3L8VpHp (SyncedReview)

Reports mentioned **Levenshtein Transformer**

> https://zd.net/3sdVEIN (ZDNET)

Reports on **Non-Autoregressive Neural Machine Translation**

> https://cnb.cx/3siXyHV (CNBC)
>
> https://bit.ly/3GqC4hw (VentureBeat)

Reports on **Learning to Translate in Real-time with Neural Machine Translation**

> https://bit.ly/3sgr5SL (Slator)

Awards & Honors

World's Top 2% Scientists (Single Year)	2023, 2022, 2021
IEEE CI HK FYP Competition and Graduate Student Paper Competition 2015/16, 2nd Runner-up	2016
HKPFS 2014/15 (Hong Kong PhD Fellowship Scheme) Award	2014
CSC Scholarship for Excellent Undergraduate Exchange Program	2012
National Endeavor Scholarship Award	2013, 2012, 2011

Publications

*=equal contribution

Recent Preprints

[1] Georgia Gabriela Sampaio, Ruixiang Zhang, Shuangfei Zhai, **Jiatao Gu**, Josh Susskind, Navdeep Jaitly, and Yizhe Zhang. "TypeScore: A Text Fidelity Metric for Text-to-Image Generative Models". In: *arXiv preprint arXiv:2411.02437* (2024).

[2] **Jiatao Gu**, Yuyang Wang, Yizhe Zhang, Qihang Zhang, Dinghuai Zhang, Navdeep Jaitly, Josh Susskind, and Shuangfei Zhai. "DART: Denoising Autoregressive Transformer for Scalable Text-to-Image Generation". In: *arXiv preprint arXiv:2410.08159* (2024).

[3] Dinghuai Zhang, Yizhe Zhang, **Jiatao Gu**, Ruixiang Zhang, Josh Susskind, Navdeep Jaitly, and Shuangfei Zhai. "Improving GFlowNets for Text-to-Image Diffusion Alignment". In: *arXiv preprint arXiv:2406.00633* (2024).

[4] Chen Wang, **Jiatao Gu**, Xiaoxiao Long, Yuan Liu, and Lingjie Liu. "GECO: Generative Image-to-3D within a SECOnd". In: *arXiv preprint arXiv:2405.20327* (2024).

[5] Ying Shen, Yizhe Zhang, Shuangfei Zhai, Lifu Huang, Joshua M Susskind, and **Jiatao Gu**. "Many-to-many Image Generation with Auto-regressive Diffusion Models". In: *arXiv preprint arXiv:2404.03109* (2024).

[6] Liangchen Song, Liangliang Cao, **Jiatao Gu**, Yifan Jiang, Junsong Yuan, and Hao Tang. "Efficient-NeRF2NeRF: Streamlining Text-Driven 3D Editing with Multiview Correspondence-Enhanced Diffusion Models". In: 2023.

[7] Samira Abnar, Omid Saremi, Laurent Dinh, Shantel Wilson, Miguel Angel Bautista, Chen Huang, Vimal Thilak, Etai Littwin, **Jiatao Gu**, Josh Susskind, et al. "Adaptivity and Modularity for Efficient Generalization Over Task Complexity". In: *arXiv preprint arXiv:2310.08866*. 2023.

Conferences

[1] **Jiatao Gu**, Ying Shen, Shuangfei Zhai, Yizhe Zhang, Navdeep Jaitly, and Joshua M Susskind. "Kaleido Diffusion: Improving Conditional Diffusion Models with Autoregressive Latent Modeling". In: *NeurIPS* (2024).

[2] Zhuofeng Wu, He Bai, Aonan Zhang, **Jiatao Gu**, VG Vydiswaran, Navdeep Jaitly, and Yizhe Zhang. "Divide-or-Conquer? Which Part Should You Distill Your LLM?" In: *EMNLP* (2024).

[3] Yizhe Zhang, He Bai, Ruixiang Zhang, **Jiatao Gu**, Shuangfei Zhai, Josh Susskind, and Navdeep Jaitly. "How Far Are We from Intelligent Visual Deductive Reasoning?" In: *COLM* (2024).

[4] **Jiatao Gu**, Shuangfei Zhai, Yizhe Zhang, Lingjie Liu, and Joshua M Susskind. "BOOT: Data-free Distillation of Denoising Diffusion Models with Bootstrapping". In: *ICML*. 2024.

[5] Tianrong Chen, **Jiatao Gu**, Laurent Dinh, Evangelos A Theodorou, Josh Susskind, and Shuangfei Zhai. "Generative Modeling with Phase Stochastic Bridges". In: *ICLR*. 2024.

[6] **Jiatao Gu**, Shuangfei Zhai, Yizhe Zhang, Josh Susskind, and Navdeep Jaitly. "Matryoshka Diffusion Models". In: *ICLR*. 2024.

[7] Jing Nathan Yan*, **Jiatao Gu***, and Alexander Rush. "Diffusion Models Without Attention". In: *CVPR*. 2024.

[8] **Jiatao Gu**, Qingzhe Gao, Shuangfei Zhai, Baoquan Chen, Lingjie Liu, and Josh Susskind. "Learning Controllable 3D Diffusion Models from Single-view Images". In: *3DV*. 2024.

[9] Yizhe Zhang, **Jiatao Gu**, Zhuofeng Wu, Shuangfei Zhai, Josh Susskind, and Navdeep Jaitly. "PLANNER: Generating Diversified Paragraph via Latent Language Diffusion Model". In: *NeurIPS*. 2023.

[10] Hansheng Chen, **Jiatao Gu**, Anpei Chen, Wei Tian, Zhuowen Tu, Lingjie Liu, and Hao Su. "Single-Stage Diffusion NeRF: A Unified Approach to 3D Generation and Reconstruction". In: *ICCV*. 2023.

[11] Shuangfei Zhai, Tatiana Likhomanenko, Etai Littwin, Dan Busbridge, Jason Ramapuram, Yizhe Zhang, **Jiatao Gu**, and Joshua M Susskind. "Stabilizing Transformer Training by Preventing Attention Entropy Collapse". In: *ICML*. 2023.

[12] **Jiatao Gu**, Alex Trevithick, Kai-En Lin, Josh Susskind, Christian Theobalt, Lingjie Liu, and Ravi Ramamoorthi. "NerfDiff: Single-image View Synthesis with NeRF-guided Distillation from 3D-aware Diffusion". In: *ICML*. 2023.

[13] **Jiatao Gu**, Shuangfei Zhai, Yizhe Zhang, Miguel Angel Bautista, and Josh Susskind. "f-dm: A multi-stage diffusion model via progressive signal transformation". In: *ICLR*. 2023.

[14] Peiye Zhuang, Samira Abnar, **Jiatao Gu**, Alex Schwing, Joshua M Susskind, and Miguel Ángel Bautista. "Diffusion probabilistic fields". In: *ICLR*. 2023.

[15] Chen Huang, Hanlin Goh, **Jiatao Gu**, and Josh Susskind. "MAST: Masked Augmentation Subspace Training for Generalizable Self-Supervised Priors". In: *ICLR*. 2023.

[16] Ann Lee, Peng-Jen Chen, Changhan Wang, **Jiatao Gu**, Xutai Ma, Adam Polyak, Yossi Adi, Qing He, Yun Tang, Juan Pino, et al. "Direct speech-to-speech translation with discrete units". In: ACL. 2022.

[17] Sravya Popuri, Peng-Jen Chen, Changhan Wang, Juan Pino, Yossi Adi, **Jiatao Gu**, Wei-Ning Hsu, and Ann Lee. "Enhanced direct speech-to-speech translation using self-supervised pre-training and data augmentation". In: *Interspeech*. 2022.

[18] Ann Lee, Hongyu Gong, Paul-Ambroise Duquenne, Holger Schwenk, Peng-Jen Chen, Changhan Wang, Sravya Popuri, Juan Pino, **Jiatao Gu**, and Wei-Ning Hsu. "Textless Speech-to-Speech Translation on Real Data". In: *NAACL*. 2022.

[19] Alexei Baevski, Wei-Ning Hsu, Qiantong Xu, Arun Babu, **Jiatao Gu**, and Michael Auli. "Data2vec: A General Framework for Self-supervised Learning in Speech, Vision and Language". In: *ICML*. 2022.

[20] Yun Tang, Hongyu Gong, Ning Dong, Changhan Wang, Wei-Ning Hsu, **Jiatao Gu**, Alexei Baevski, Xian Li, Abdelrahman Mohamed, Michael Auli, et al. "Unified speech-text pre-training for speech translation and recognition". In: *ACL*. 2022.

[21] Zhuofeng Wu, Sinong Wang, **Jiatao Gu**, Rui Hou, Yuxiao Dong, VG Vydiswaran, and Hao Ma. "IDPG: An instance-dependent prompt generation method". In: *ACL*. 2022.

[22] Khalil Mrini, Shaoliang Nie, **Jiatao Gu**, Sinong Wang, Maziar Sanjabi, and Hamed Firooz. "Detection, disambiguation, re-ranking: autoregressive entity linking as a multi-task problem". In: *ACL*. 2022.

[23] **Jiatao Gu** and Xu Tan. "Non-autoregressive sequence generation". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*. 2022, pp. 21–27.

[24] **Jiatao Gu**, Lingjie Liu, Peng Wang, and Christian Theobalt. "StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis". In: ICLR. 2022.

[25] Jiajun Shen, Peng-Jen Chen, Matt Le, Junxian He, **Jiatao Gu**, Myle Ott, Michael Auli, and Marc'Aurelio Ranzato. "The source-target domain mismatch problem in machine translation". In: EACL. 2021.

[26] Lingjie Liu, Marc Habermann, Viktor Rudnev, Kripasindhu Sarkar, **Jiatao Gu**, and Christian Theobalt. "Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control". In: *SIGGRAPH ASIA*. 2021.

[27] Lior Yariv, **Jiatao Gu**, Yoni Kasten, and Yaron Lipman. "Volume rendering of neural implicit surfaces". In: *NeurIPS*. (oral). 2021.

[28] Changhan Wang, Wei-Ning Hsu, Yossi Adi, Adam Polyak, Ann Lee, Peng-Jen Chen, **Jiatao Gu**, and Juan Pino. "fairseq Sˆ 2: A Scalable and Integrable Speech Synthesis Toolkit". In: EMNLP Demo track. 2021.

[29] Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, **Jiatao Gu**, and Angela Fan. "Multilingual translation with extensible multilingual pretraining and finetuning". In: ACL Findings. 2021.

[30] Chunting Zhou, **Jiatao Gu**, Mona Diab, Paco Guzman, Luke Zettlemoyer, and Marjan Ghazvininejad. "Detecting Hallucinated Content in Conditional Neural Sequence Generation". In: ACL Findings. 2021.

[31] **Jiatao Gu*** and Xiang Kong*. "Fully Non-autoregressive Neural Machine Translation: Tricks of the Trade". In: ACL Findings. 2021.

[32] Xiang Kong, Adithya Renduchintala, James Cross, Yuqing Tang, **Jiatao Gu**, and Xian Li. "Multilingual Neural Machine Translation with Deep Encoder and Multiple Shallow Decoders". In: *EACL*. 2021, pp. 1613–1624.

[33] Changhan Wang, Kyunghyun Cho, and **Jiatao Gu**. "Neural machine translation with byte-level subwords". In: *AAAI*. Vol. 34. 05. 2020, pp. 9154–9160.

[34] Junxian He*, **Jiatao Gu***, Jiajun Shen, and Marc'Aurelio Ranzato. "Revisiting self-training for neural sequence generation". In: *ICLR*. 2020.

[35] Maha Elbayad, **Jiatao Gu**, Edouard Grave, and Michael Auli. "Depth-adaptive transformer". In: *ICLR*. 2020.

[36] Chunting Zhou*, **Jiatao Gu***, and Graham Neubig. "Understanding knowledge distillation in non-autoregressive machine translation". In: *ICLR*. 2020.

[37] Xutai Ma, Juan Pino, James Cross, Liezl Puzon, and **Jiatao Gu**. "Monotonic multihead attention". In: *ICLR*. 2020.

[38] Changhan Wang, Juan Pino, Anne Wu, and **Jiatao Gu**. "Covost: A diverse multilingual speech-to-text translation corpus". In: *LREC*. 2020.

[39] Changhan Wang, Juan Pino, and **Jiatao Gu**. "Improving Cross-Lingual Transfer Learning for End-to-End Speech Recognition with Speech Translation". In: *Interspeech*. 2020.

[40] Chau Tran, Yuqing Tang, Xian Li, and **Jiatao Gu**. "Cross-lingual retrieval for iterative self-supervised training". In: *NeurIPS*. 2020.

[41] Ebrahim Ansari, Nguyen Bach, Ondřej Bojar, Roldano Cattoni, Fahim Dalvi, Nadir Durrani, Marcello Federico, Christian Federmann, **Jiatao Gu**, Fei Huang, et al. "Findings of the IWSLT 2020 Evaluation Campaign". In: Proceedings of the 17th International Conference on Spoken Language Translation. 2020, pp. 1–34.

[42] Anne Wu, Changhan Wang, Juan Pino, and **Jiatao Gu**. "Self-supervised representations improve end-to-end speech translation". In: *Interspeech*. 2020.

[43] Arya D McCarthy, Xian Li, **Jiatao Gu**, and Ning Dong. "Addressing Posterior Collapse with Mutual Information for Improved Variational Neural Machine Translation". In: *ACL*. 2020, pp. 8512–8525.

[44] Lingjie Liu*, **Jiatao Gu***, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. "Neural Sparse Voxel Fields". In: *NeurIPS*. Vol. 33. (spotlight). 2020.

[45] Saining Xie, **Jiatao Gu**, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. "PointContrast: Unsupervised pre-training for 3D point cloud understanding". In: *ECCV*. Springer, Cham. 2020, pp. 574–591.

[46] Xutai Ma, Mohammad Javad Dousti, Changhan Wang, **Jiatao Gu**, and Juan Pino. "Simuleval: An evaluation toolkit for simultaneous translation". In: *EMNLP Demo Track*. 2020.

[47] Hang Le, Juan Pino, Changhan Wang, **Jiatao Gu**, Didier Schwab, and Laurent Besacier. "Dual-decoder Transformer for Joint Automatic Speech Recognition and Multilingual Speech Translation". In: COLING. 2020.

[48] Jungo Kasai, James Cross, Marjan Ghazvininejad, and **Jiatao Gu**. "Non-autoregressive machine translation with disentangled context transformer". In: *ICML*. PMLR. 2020, pp. 5144–5155.

[49] Peng-Jen Chen, Ann Lee, Changhan Wang, Naman Goyal, Angela Fan, Mary Williamson, and **Jiatao Gu**. "Facebook AI's WMT20 News Translation Task Submission". In: WMT. 2020.

[50] **Jiatao Gu***, Yong Wang*, Kyunghyun Cho, and Victor OK Li. "Improved Zero-shot Neural Machine Translation via Ignoring Spurious Correlations". In: *ACL*. 2019.

[51] **Jiatao Gu**, Changhan Wang, and Jake Zhao. "Levenshtein Transformer". In: *NeurIPS*. 2019.

[52] Changhan Wang, Anirudh Jain, Danlu Chen, and **Jiatao Gu**. "VizSeq: A Visual Analysis Toolkit for Text Generation Tasks". In: *EMNLP Demo Track*. 2019.

[53] Juan Pino, Liezl Puzon, **Jiatao Gu**, Xutai Ma, Arya D McCarthy, and Deepak Gopinath. "Harnessing indirect training data for end-to-end automatic speech translation: Tricks of the trade". In: IWSLT. 2019.

[54] **Jiatao Gu**, Yong Wang, Kyunghyun Cho, and Victor OK Li. "Search Engine Guided Non-Parametric Neural Machine Translation". In: *AAAI*. 2018.

[55] **Jiatao Gu**, Daniel Jiwoong Im, and Victor OK Li. "Neural machine translation with gumbel-greedy decoding". In: *AAAI*. 2018.

[56] **Jiatao Gu**, James Bradbury, Caiming Xiong, Victor OK Li, and Richard Socher. "Non-Autoregressive Neural Machine Translation". In: *ICLR*. 2018.

[57] **Jiatao Gu**, Hany Hassan, Jacob Devlin, and Victor OK Li. "Universal Neural Machine Translation for Extremely Low Resource Languages". In: *NAACL*. 2018.

[58] **Jiatao Gu**, Yong Wang, Yun Chen, Kyunghyun Cho, and Victor OK Li. "Meta-learning for low-resource neural machine translation". In: *EMNLP*. 2018.

[59] **Jiatao Gu**, Graham Neubig, Kyunghyun Cho, and Victor OK Li. "Learning to translate in real-time with neural machine translation". In: *EACL*. 2017.

[60] **Jiatao Gu**, Kyunghyun Cho, and Victor OK Li. "Trainable greedy decoding for neural machine translation". In: *EMNLP*. 2017.

[61] Victor O. K. Li, Jacqueline C. K. Lam, Yun Chen, and **Jiatao Gu**. "Deep Learning Model to Estimate Air Pollution Using M-BP to Fill in Missing Proxy Urban Data". In: *GLOBECOM*. 2017.

[62] **Jiatao Gu**, Zhengdong Lu, Hang Li, and Victor OK Li. "Incorporating copying mechanism in sequence-to-sequence learning". In: *ACL*. 2016.

[63] **Jiatao Gu** and Victor OK Li. "Efficient Learning for Undirected Topic Models". In: *ACL Short*. 2015.

JOURNALS

[1] Peng Wang, Yuan Liu, Guying Lin, **Jiatao Gu**, Lingjie Liu, Taku Komura, and Wenping Wang. "ProLiF: Progressively-connected Light Field network for efficient view synthesis". In: *Computers & Graphics* 120 (2024), p. 103913.

[2] Yinhan Liu*, **Jiatao Gu***, Naman Goyal*, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. "Multilingual denoising pre-training for neural machine translation". In: (2020).

[3] **Jiatao Gu**, Qi Liu, and Kyunghyun Cho. "Insertion-based decoding with automatically inferred generation order". In: (2019).

[4] JQ James, Wen Yu, and **Jiatao Gu**. "Online vehicle routing with neural combinatorial optimization and deep reinforcement learning". In: 20.10 (2019), pp. 3806–3817.

[5] James Jian Qiao Yu and **Jiatao Gu**. "Real-time traffic speed estimation with graph convolutional generative autoencoder". In: 20.10 (2019), pp. 3940–3951.

[6] JQ James, David J Hill, Albert YS Lam, **Jiatao Gu**, and Victor OK Li. "Intelligent time-adaptive transient stability assessment system". In: 33.1 (2017), pp. 1049–1058.

WORKSHOPS/UNPUBLISHED

[1] Peng Wang, Yuan Liu, Guying Lin, **Jiatao Gu**, Lingjie Liu, Taku Komura, and Wenping Wang. "Progressively-connected Light Field Network for Efficient View Synthesis". In: *arXiv preprint arXiv:2207.04465*. 2022.

[2] Zhuofeng Wu, Sinong Wang, **Jiatao Gu**, Madian Khabsa, Fei Sun, and Hao Ma. "CLEAR: Contrastive Learning for Sentence Representation". In: arXiv preprint arXiv:2012.15466. 2020.

[3] Yong Wang, Xiao-Ming Wu, Qimai Li, **Jiatao Gu**, Wangmeng Xiang, Lei Zhang, and Victor OK Li. "Large Margin Meta-Learning for Few-Shot Classification". In: NeurIPS Meta-learning Workshop. 2018.

[4] **Jiatao Gu**, Baotian Hu, Zhengdong Lu, Hang Li, and Victor OK Li. "Guided Sequence-to-Sequence Learning with External Rule Memory". In: ICLR Workshop. 2016.