Yongqiang Wang

Last update on August 18, 2023

yongqiang.wang@outlook.com • +1-425-435-0511
yongqiang.seagull (*Skype*) • http://www.linkedin.com/in/yongqiang-wang (*Linkedin*)
New York • N. Y. • U.S.A.

Experience

Google — New York, N.Y., U.S.A.
Staff Research Scientist — *Nov. 2020 – Now*

- Research, Development and Deployment of Foundation Speech Models:
  - Develop a novel attention mechanism to stabilize learning large model on long-form speech;
  - Lead the development of Universal Speech Models for 100+ languages; this model improves over prior long-form speech recognition systems by over 50%, and runs at 800x faster than real time;
  - Deploy Universal Speech Models for Youtube and Google Cloud Speech API;
  - Lead the development of an ASR inference platform using Universal Speech Models, which is used widely cross Google.

Facebook — Seattle, W.A., U.S.A.
Staff Research Scientist — *Jan. 2017 – Nov. 2020*

- Lead the development of core ASR technology for speech products and services across Facebook;
- Build a unified platform for research, development and deployment for the next generation speech technology:
  - Support ASR, TTS, speech translation, speaker recognition and many other areas;
  - Widely adopted across Facebook: it consumes over 0.2M GPU hours daily, and supports multiple critical product delivery;
- Build an ASR modeling team ($\sim$ 25 people) from ground up.

Microsoft — Bellevue, W.A., U.S.A.
Speech Scientist — *Feb. 2014 – Sept. 2016*

- Develop deep Learning (DL)-based acoustic models on devices;
- Large scale distributed training of deep learning models;
- Key contributors to the Microsoft CNTK project since early 2015.

Education

Cambridge University — Cambridge, U.K.
PhD in Information Engineering — *2009 – 2014*

The University of Hong Kong — Hong Kong, China
Master in Computer Science — *2006 – 2009*

University of Science and Technology of China — Hefei, China
Bachelor in Electrical Engineering — *2002 – 2006*

Selected Publications

1. P. K. Rubenstein, C. Asawaroengchai, D. D. Nguyen, Ankur Bapna, et al. AudioPaLM: A Large Language Model That Can Speak and Listen. *arXiv preprint arXiv:2306.12925*, 2023

2. Y. Zhang, W. Han, J. Qin, Y. Wang, et al. Google USM: Scaling automatic speech recognition beyond 100 languages. *arXiv preprint arXiv:2303.01037*, 2023

3. Y. Wang, Z. Chen, C.-J. Zheng, Y. Zhang, W. Han, and P. Haghani. Accelerating RNN-t training and inference using CTC guidance. In *ICASSP*, pages 1–5, 2023

4. Y. Shi, Y. Wang, C. Wu, et al. Emformer: Efficient memory transformer based acoustic model for low latency streaming speech recognition. In *ICASSP*, pages 6783–6787, 2021
5. C. Wu, Y. Wang, Y. Shi, C.-F. Yeh, and F. Zhang. Streaming transformer-based acoustic models using self-attention with augmented memory. *arXiv preprint arXiv:2005.08042*, 2020
6. Y.-Q. Wang, A. Mohamed, D. Le, et al. Transformer-based acoustic modeling for hybrid speech recognition. *arXiv preprint arXiv:1910.09799*, 2019.
7. C.-F. Yeh, J. Mahadeokar, K. Kalgaonkar, Y.-Q. Wang, D. Le, M. Jain, K. Schubert, C. Fuegen, and M. Seltzer. Transformer-transducer: End-to-end speech recognition with self-attention. *arXiv preprint arXiv:1910.12977*, 2019.
8. A. Tjandra, C. Liu, F. Zhang, X. Zhang, Y.-Q. Wang, G. Synnaeve, S. Nakamura, and G. Zweig. Deja-vu: Double feature presentation in deep transformer networks. *Proc. International Conference on Acoustic, Speech, and Signal Processing* (*ICASSP*), 2020.
9. K. Singh, D. Okhonko, J. Liu, Y.-Q. Wang, F. Zhang, R. Girshick, S. Edunov, F. Peng, Y. Saraf, G. Zweig, et al. Training asr models by generation of contextual information. *arXiv preprint arXiv:1910.12367*, 2019.
10. Z. Chen, M. Jain, Y.-Q. Wang, M. Seltzer, and C. Fuegen. Joint grapheme and phoneme embeddings for contextual end-to-end asr. *Proc. Annual Conference of the International Speech Communication* (*Interspeech*), pages 3490–3494, 2019.
11. Z. Chen, M. Jain, Y.-Q. Wang, M. Seltzer, and C. Fuegen. End-to-end contextual speech recognition using class language models and a token passing decoder. In *Proc. International Conference on Acoustic, Speech, and Signal Processing* (*ICASSP*), pages 6186–6190, 2019.
12. D. Serdyuk, Y.-Q. Wang, C. Fuegen, A. Kumar, B. Liu, and Y. Bengio. Towards end-to-end spoken language understanding. In *Proc. International Conference on Acoustic, Speech, and Signal Processing* (*ICASSP*), pages 5754–5758, 2018.
13. X. Chen, X. Liu, Y.-Q. Wang, M. J. F. Gales, and P. C. Woodland. Efficient training and evaluation of recurrent neural network language models for automatic speech recognition. *IEEE Transactions on Audio, Speech and Language Processing* (*ASLP*), August 2016.
14. X. Liu, X. Chen, Y.-Q. Wang, M. J. F. Gales, and P. C. Woodland. Two efficient lattice rescoring methods using recurrent neural network language models. *IEEE Transactions on Audio, Speech and Language Processing* (*ASLP*), 24(8):1438–1449, August 2016.
15. Y. Huang, Y.-Q. Wang, and Y. Gong. Semi-supervised training in deep learning acoustic models. In *Proc. Annual Conference of the International Speech Communication* (*Interspeech*), 2016.
16. Y.-J. Miao, J. Li, Y.-Q. Wang, S. Zhang, and Y. Gong. Simplifying long short-term memory acoustic models for fast training and decoding. In *Proc. International Conference on Acoustic, Speech, and Signal Processing* (*ICASSP*), 2016.
17. C. Liu, Y.-Q. Wang, K. Kumar, and Y. Gong. Investigations on speaker adaptation of lstm rnn models for speech recognition. In *Proc. International Conference on Acoustic, Speech, and Signal Processing* (*ICASSP*), 2016.
18. Y.-Q. Wang, J. Li, and Y. Gong. Small-footprint high-performance deep Neural network-based speech recognition using split-VQ. In *Proc. International Conference on Acoustic, Speech, and Signal Processing* (*ICASSP*), 2015.
19. P. Karanasou, Y.-Q. Wang, M. J. F. Gales, and P. C. Woodland. Adaptation of deep neural network acoustic models using factorised i-Vectors. In *Proc. Annual Conference of the International Speech Communication* (*Interspeech*), 2014.
20. M. Seltzer, D. Yu, and Y.-Q. Wang. An investigation of deep neural networks for noise robust speech recognition. In *Proc. International Conference on Acoustic, Speech, and Signal Processing* (*ICASSP*), 2013.
21. Y.-Q. Wang and M. J. F. Gales. Speaker and noise factorisation for robust speech recognition. *IEEE Transactions on Audio, Speech and Language Processing* (*ASLP*), 20(7), 2012.