Po-Yao (Bernie) Huang

Mail: berniebear@gmail.com; Google Scholar: https://tinyurl.com/p6n7fhmb; Website: http://www.cs.cmu.edu/~poyaoh
FAIR Labs, Meta AI

EXPERIENCE

Meta AI, Menlo Park, USA
- Senior Research Scientist (FAIR Labs) 2021.8 ~ present
- Research Intern 2020.5 ~ 2021.5

Microsoft, Redmond, USA
- Research Intern 2017.6 ~ 2017.8

MediaTek, Taipei, Taiwan
- Senior Software Engineer 2010.9 ~ 2014.5

EDUCATION

Carnegie Mellon University, Pittsburgh, USA
- Ph.D. in Language and Information Technologies 2016.8 ~ 2021.7
- M.S. in Language Technologies 2014.8 ~ 2016.7
  - Rank: 1/40, Top 1%, GPA: 4.21 /4.33
  - Advisor: Alexander G. Hauptmann

National Taiwan University, Taipei, Taiwan
- M.S. in Computer Engineering 2003.9 ~ 2007.6
  - Rank 3/127, Top 2%, GPA: 4.00/4.00
- B.S. in Electrical Engineering 2003.9 ~ 2007.6
  - GPA: 3.87/4.00

PUBLICATION (Selected)

- "Simple MViT: A Hierarchical Vision Transformer without the Bells-and-Whistles," C. Ryali, Y-T. Hu, D. Bolya, C. Wei, H. Fan, Po-Yao Huang, V. Aggarwal, A. Chowdhury, O. Poursaeed, J. Hoffman, J. Malik, Y. Li, C. Feichtenhofer. ICML 2023
- "Masked Autoencoders that Listen," Po-Yao Huang, H. Xu, J. Li, A. Baevski, M. Auli, W. Galuba, F. Metze, C. Feichtenhofer. NeurIPS 22.
- "Video Pivoting Unsupervised Multi-Modal Machine Translation," M. Li, Po-Yao Huang, X. Chang, J. Hu, Y. Yang, A. Hauptmann. IEEE T-PAMI 22.
- "On adversarial robustness of large-scale audio visual learning," J. Li, S. Qu, X. Li, Po-Yao Huang, F. Metze. ICASSP 22 (Best Student Paper).
- "Multilingual Multimodal Pre-training for Zero-Shot Cross-Lingual Transfer of Vision-Language Models," Po-Yao Huang*, M. Patrick*, J. Hu, G. Neubig, F. Metze, A. Hauptmann, NAACL 21.
- "Videoclip: Contrastive pre-training for zero-shot video-text understanding," H. Xu, G. Ghosh, Po-Yao Huang, D. Okhonko, A. Aghajanyan, F. Metze, L. Zettlemoyer, C. Feichtenhofer. EMNLP 21.
- "Support-set bottlenecks for video-text representation learning," M. Patrick*, Po-Yao Huang*, Y. Asano*, F. Metze, A. Hauptmann, J. Henriques, A. Vedaldi. ICLR 21.
- "Self-supervised Deep Correlation Tracking," D. Yuan, X. Chang, Po-Yao Huang, Q. Liu, Z. He, IEEE TIP 21.
- "Space-time crop & attend: Improving cross-modal video representation learning," M. Patrick*, Po-Yao Huang*, I. Misra, F. Metze, A. Vedaldi, Y. Asano, J. F Henriques. ICCV 21.
- "Argus: Efficient activity detection system for extended video analysis," W. Liu*, G. Kang*, Po-Yao Huang*, et al. WACV 20.
- "Unsupervised Multimodal Neural Machine Translation with Pseudo Visual Pivoting," Po-Yao Huang, J. Hu, X. Chang, A. Hauptmann. ACL 20.

- "Multi-Head Attention with Diversity for Learning Grounded Multilingual Multimodal Representations," Po-Yao Huang, X. Chang, A. Hauptmann. EMNLP 19.
- "Annotation Efficient Cross-Modal Retrieval with Adversarial Attentive Alignment," Po-Yao Huang, G. Kang, W. Liu, X, Chang, A. Hauptmann. ACM MM 19.
- "RCAA: Relational Context-Aware Agents for Person Search," X. Chang, Po-Yao Huang, Y. Shen, X. Liang, Y. Yang, A. Hauptmann. ECCV 18.
- "Multimodal Filtering of Social Media for Temporal Monitoring and Event Analysis," Po-Yao Huang, J.W. Liang, J. Lamare, A. Hauptmann. ICMR 18.
- "Attention-based Multimodal Neural Machine Translation," Po-Yao Huang, F. Liu, S. Shiang, C. Dyer. WMT 16.
- "Entity Hierarchy Embedding," Z. Hu, Po-Yao Huang, Y. Deng, Y. Gao, E. Xing. ACL 15.

PREPRINTS

- "Dinov2: Learning robust visual features without supervision," M. Oquab et al.
- "Diffusion Models as Masked Autoencoders," C. Wei, K. Mangalam, Po-Yao Huang, Y. Li, H. Fan, H. Xu, H. Wang, C. Xie, A. Yuille, C. Feichtenhofer.
- "MAViL: Masked Audio-Video Learners," Po-Yao Huang, V. Sharma, H. Xu, C. Ryali, H. Fan, Y. Li, S. Li, G. Ghosh, J. Malik, C. Feichtenhofer.
- "CiT: Curation in Training for Effective Vision-Language Data," H. Xu, S. Xie, Po-Yao Huang, L. Yu, R. Howes, G. Ghosh, L. Zettlemoyer, C. Feichtenhofer.
- "Cm3: A causal masked multimodal model of the internet," A. Aghajanyan, Po-Yao Huang, C. Ross, V. Karpukhin, H. Xu, N. Goyal, D. Okhonko, M. Joshi, G. Ghosh, M. Lewis, L. Zettlemoyer.

PATENT

- A. Hauptmann, Po-Yao Huang, P. Sahin, "Multi-model Monitoring and Coaching System to Promote Proper Asthma Inhaler Technique" U.S. Patent, 62/708,345, Aug 2019
- A. Hauptmann, L. Jiang and Po-Yao Huang, "Large-Scale Video Content Retrieval System Through Text Query." U.S. Patent, 15/769,233, Oct 2018

PROJECT

- [FAIR Labs] Multimodal Large Language Model: Develop the generative causal masked multimodal models (CM3) which are trained on large-scale structured multi-modal documents containing both web-crawled text and image tokens.
- [FAIR Labs] Multimodal Self-Supervised Learning: Self-supervised representation learning from audio spectrograms, video, text. Recent project Audio-MAE sets new state-of-the-art performance on six audio and speech classification tasks. Demo, IEEE Specrtum News coverage.
- [CMU] Recognition of Activities in Extended Video (ActEV/DIVA): Developing video analysis system detecting activities in surveillance scenarios. Our system achieves state-of-the-art surveillance video analysis in various challenges.

SERVICE

- Reviewer: NAACL, EMNLP, ACL, EMNLP, ICLR, ACM MM
- Area Chair: ACM MM, NAACL

HONOR & AWARD

- 1st Place in TRECVID-ActEV challenge 2020
- 1st Place in TRECVID-ActEV challenge 2019
- 1st Place in TREC-Incident Streams challenge 2019
- 2nd Place in Kinetics-800 Action Recognition Challenge in ActivityNet 2019
- 1st Place in ActEV 2019 challenge in ActivityNet 2019
- 2nd Place in TRECVID-AVS (Ad-hoc Video Search) challenge 2018
- 5th Place in Moment in Time Challenge 2018
- 2nd Place in Google Youtube 8M Challenge 2017
- Siebel Scholar 2016 (Top CS/Business students in top graduate schools worldwide)