

Education

- University of California, San Diego** San Diego, CA
• *Ph.D in Electrical Engineering; GPA:4.0* Sept 2018
Advisor: Bhaskar D. Rao, co-advisor: Truong Q. Nguyen
- University of Illinois at Urbana-Champaign** Urbana-Champaign, IL
• *M.S. in Electrical Engineering; GPA:3.78* May 2014
Advisor: Pierre Moulin
- University of Illinois at Urbana-Champaign** Urbana-Champaign, IL
• *B.S. in Electrical Engineering; GPA:3.90* May 2012
James Scholar, Highest Honors

Experience

- Meta** Remote
• *AI Research Scientist* June 2022 – Present
 - AutoML and neural architecture search for ads recommendation system
- ARM Research** Boston, MA
• *Staff Research Engineer* April 2021 – June 2022
 - Machine learning research group
 - Neural architecture search for hardware-aware, efficient neural networks
 - Differentiable neural architecture search, Bayesian optimization, unstructured pruning, channel pruning, quantization, compression
- ARM Research** Boston, MA
• *Senior Research Engineer* Sept 2018 – April 2021
 - Machine learning research group
- Samsung Research** San Diego, CA
• *Intern* June 2017 – Sept 2017
 - Deep learning research group
- Qualcomm** San Diego, CA
• *Intern* May 2015 – Aug 2015
 - Developed continuous multi-modal authentication system for verifying mobile user's identity
- Qualcomm** San Diego, CA
• *Intern* May 2013 – Sept 2014
 - Developed real-time, fixed point C implementation of Fast Stereo Independent Vector Analysis
- Qualcomm** San Diego, CA
• *Intern* Jun 2012 – Aug 2012
 - Developed novel voice activity detector using non-negative matrix factorization
- Cisco** San Jose, CA
• *Intern* Jun 2011 – Aug 2011
 - Implemented testing framework for NX-OS
- ComEd** Libertyville, IL
• *Intern* Jun 2010 – Aug 2010
 - Worked with Transmission and Substation Department in the Testing Group

Patents

- M. Haddon, **I. Fedorov**, R. Jeyapaul, P. N. Whatmough, Z. Liu, “Error detection,” *US patent application*.
- H. Tann., R. Navarro, **I. Fedorov**, C. Zhou, P. Whatmough, M. Mattina, “System, Devices and/or Processes for Defining a Search Space for Neural Network Processing Device Architectures,” *US patent application*.
- **I. Fedorov**, P. Whatmough, “Neural network system and training method,” *US patent application*.
- **I. Fedorov**, R. Matas, C. Zhou, H. Tann, P. Whatmough, M. Mattina, “A unified neural network optimization framework,” *US patent application*.
- M. El-Khamy **I. Fedorov**, J. Lee, “Image denoising neural network architecture and method of training the same,” *US patent*, 2020.

Publications (by topic)

Neural Architecture Search

- Y. Chai, D. Tripathy, C. Zhou, D. Gope, **I. Fedorov**, R. Matas, D. Brooks, G. Wei, P. Whatmough, “PerfSAGE: Generalized Inference Performance Predictor for Arbitrary Deep Learning Models on Edge Devices,” *ArXiv*, 2023.
- A. Kag, **I. Fedorov**, A. Gangrade, P.N. Whatmough, V. Saligrama, “Efficient Edge Inference by Selective Query,” *ICLR*, 2023.
- **I. Fedorov**, R. Matas, H. Tann, C. Zhou, M. Mattina, P.N. Whatmough, “UDC: Unified DNAs for Compressible TinyML Models,” *NeurIPS*, 2022.
- K. Bhardwaj, J. Ward, C. Tung, D. Gope, L. Meng, **I. Fedorov**, A. Chalfin, P. Whatmough, D. Loh, “Restructurable Activation Networks,” *ArXiv*, 2022.
- A. Kag, **I. Fedorov**, A. Gangrade, P.N. Whatmough, V. Saligrama, “Achieving High TinyML Accuracy through Selective Cloud Interactions,” *ICML DyNN workshop*, 2022.
- C. Banbury*, C. Zhou*, **I. Fedorov***, R.M. Navarro, U. Thakker, D. Gope, V.J. Reddi, M. Mattina, P.N. Whatmough, “MicroNets: Neural Network Architectures for Deploying TinyML Applications on Commodity Microcontrollers,” *MLSys*, 2021.
- **I. Fedorov**, M. Stamenovic, C. Jensen, L. Yang, A. Mandell, Y. Gan, M. Mattina, P.N. Whatmough, “TinyLSTMs: Efficient Neural Speech Enhancement for Hearing Aids,” *INTERSPEECH*, 2020.
- S. Sandha, M. Aggarwal, **I. Fedorov**, M. Srivastava, “Mango: A Python Library for Parallel Hyperparameter Tuning,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2020.
- **I. Fedorov**, R.P. Adams, M. Mattina, P.N. Whatmough, “SpArSe: Sparse Architecture Search for CNNs on Resource-Constrained Microcontrollers,” *Proc. of the Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

RNN Compression

- U. Thakker, **I. Fedorov**, J. Beu, D. Gope, C. Zhou, G. Dasika M. Mattina, “Pushing the limits of RNN Compression,” *NeurIPS Workshop on Energy Efficient Machine Learning and Cognitive Computing*, 2019.

Multimodal Dictionary Learning

- **I. Fedorov**, B.D. Rao, “Multimodal Sparse Bayesian Dictionary Learning,” *arXiv preprint*, 2018.

- **I. Fedorov**, B.D. Rao, T.Q. Nguyen, “Multimodal Sparse Bayesian Dictionary Learning Applied to Multimodal Data Classification,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

Sparsifying Deep Neural Networks

- C. Lee, **I. Fedorov**, B.D. Rao, H. Garudadri, “SSGD: Sparsity-promoting Stochastic Gradient Descent Algorithm for Unbiased DNN Pruning,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2020.
- **I. Fedorov**, B.D. Rao, “Sparsifying Deep Neural Networks,” arXiv preprint, 2018.

Non-negative Matrix Factorization

- **I. Fedorov**, A. Nalci, R. Giri, B.D. Rao, T.Q. Nguyen, H. Garudadri, “A Unified Framework for Sparse Non-Negative Least Squares using Multiplicative Updates and the Non-Negative Matrix Factorization Problem,” *Signal Processing*, Volume 146, May 2018, Pages 79-91, ISSN 0167-1648.
- A. Nalci, **I. Fedorov**, M. Al-Shoukairi, T. T. Liu, B.D. Rao. “Rectified Gaussian Scale Mixtures and the Sparse Non-Negative Least Squares Problem,” *IEEE Transactions on Signal Processing*, vol. 66, no. 12, pp. 3124-3139, June 2018.

Robust Sparse Signal Recovery & Face Recognition

- **I. Fedorov**, R. Giri, B.D. Rao, T.Q. Nguyen, “Relevance Vector Machine: A Novel Person Re-Identification Framework,” *arXiv preprint arXiv:1703.10645*, 2017.
- **I. Fedorov**, R. Giri, B.D. Rao, T.Q. Nguyen, “Robust Bayesian Method for Simultaneous Block Sparse Signal Recovery with Applications to Face Recognition,” *IEEE International Conference on Image Processing (ICIP)*, 2016.

Single Photon Emission Computed Tomography

- **I. Fedorov**, S. Obrzut, B. Song, B.D. Rao, “SPECT Image Reconstruction under Imaging Time Constraints,” *51st Asilomar Conference on Signals, Systems, and Computers*, 2017.
- **I. Fedorov**, B. Song, B.D. Rao, I. Levitan, S. Obrzut, “Total Variation Regularization in I-123 Ioflupane SPECT Reconstruction,” *Journal of Nuclear Medicine*, 2017.

Action Recognition

- **I. Fedorov**, “Kinect depth video compression for action recognition,” *Master’s thesis*, 2014.
- A. Khosrowpour, **I. Fedorov**, A. Holynski, J.C. Niebles, and M. Golparvar-Fard, “Automated Worker Activity Analysis in Indoor Environments for Direct-Work Rate Improvement from long sequences of RGB-D Images,” *Construction Research Congress: Construction in a Global Network*, 2014.

Miscellaneous

- P.S. Shenoy, **I. Fedorov**, T. Neyens, P.T. Krein, “Power delivery for series connected voltage domains in digital circuits,” *International Conference on Energy Aware Computing*, 2011.

Skills

Python, Tensorflow, Pytorch, Matlab, C/C++, LaTeX, Fluent in Russian

Teaching

WES 267: Intro to Digital Signal Processing, UCSD
 ECE 161B: Digital Signal Processing, UCSD
 ECE 445: Senior Design, UIUC

Honors and Activites

ARCS Fellowship, 2015-2018

ECE Departmental Fellowship, UCSD, 2014

Jules D. Falzer Scholarship for outstanding scholastic record, UIUC, 2012