VIET ANH TRINH

vtrinh@gradcenter.cuny.edu | 551-214-6767

Education

2016 - 2021 Ph.D. in Computer Science, The Graduate School and University Center, The City University of New York (CUNY), US

- Start date: Aug/25/2016. Date of completion: Dec/02/2021. Degree received date: Feb/01/2022
- Advisor: Professor Michael Mandel
- GPA: 3.916

2016 - 2021 Master of Philosophy in Computer Science, The Graduate School and University Center, The City University of New York (CUNY), US

- Start date: Aug/25/2016. Date of completion: Feb/17/2021. Degree received date: June/03/2021

2003 - 2008 B.S. in Electronics and Telecommunications, Hanoi University of Science and Technology, Vietnam

- Start date: Aug/15/2003. Date of completion: May/31/2008. Degree received date: July/04/2008.
- Top 1% of the Electronics and Telecommunications Department

Publications

V. A. Trinh and S. Braun, "Unsupervised speech enhancement with speech recognition embedding and disentanglement losses," *ICASSP*, 2022.

V. A. Trinh, H. S. Kavaki, and M. I. Mandel, "Importantaug: A data augmentation agent for speech," *ICASSP*, 2022.

V. A. Trinh and A. Rozovskaya, "New Dataset and Strong Baselines for the Grammatical Error Correction of Russian," in *Findings of ACL*, 2021.

V. A. Trinh and M. I. Mandel, "Directly comparing the listening strategies of humans and machines," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 312–323, 2021.

V. A. Trinh and M. I. Mandel, "Large Scale Evaluation of Importance Maps in Automatic Speech Recognition," in *Proc. Interspeech 2020*, pp. 1166–1170, 2020.

V. A. Trinh, B. McFee, and M. I. Mandel, "Bubble cooperative networks for identifying important speech cues," in *Proc. Interspeech 2018*, pp. 1616–1620, 2018.

A. R. Syed, V. A. Trinh, and M. I. Mandel, "Concatenative resynthesis with improved training signals for speech enhancement," in *Proc. Interspeech 2018*, pp. 1195–1199, 2018.

Work experience

2021-current **Amazon**, Alexa Speech Acoustic Modeling team, Applied Scientist, US

- Start date: Dec/06/2021
- Project: Automatic speech recognition, speech classification.

2021 **Microsoft**, Audio and Acoustics Research Group, Research Intern, US

- Start date: June/01/2021. End date: Aug/20/2021
- Address: Remote internship due to Covid-19. Microsoft physical address: One Microsoft Way, Redmond, WA, 98052.
- Project: Unsupervised speech enhancement with automatic speech recognition embedding and disentanglement losses
- Responsibilities:
  - Researched new techniques and methods to mitigate the trade-off between automatic speech recognition and speech enhancement
  - Implemented the proposed methods on large-scale computing resources
- Contributions:
  - Proposed a new unsupervised loss function that combined speech recognition embedding and disentanglement loss.
- Achievements:

- The proposed speech recognition embedding and disentanglement loss term helped to separate noise from speech while maintaining speech recognition performance
- The proposed loss function improved speech recognition performance when using speech enhancement as front-end.
- Published a paper at The International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2022:

- Techniques and methodologies: Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), Wav2Vec, Short-time Fourier Transform (STFT), AdamW optimization, unsupervised training, skip connections.
- Tools: Python, Pytorch.

2020 **Amazon**, Alexa Speech Acoustic Modeling team, Applied Scientist Intern, US

- Start date: May/26/2020. End date: Aug/21/2020
- Address: Remote internship due to Covid-19.
- Project: Improving automatic speech recognition (ASR)
- Responsibilities:
  - Enhanced Alexa ASR acoustic modeling performance
  - Implemented the proposed method on large-scale computing resources from Amazon Web Services
- Achievements: Reduced ASR error rate

2019 **Amazon**, Alexa Speech Acoustic Modeling team, Applied Scientist Intern, US

- Start date: May/28/2019. End date: Aug/23/2019
- Address: 300 Pine Street, Seattle, WA, 98101
- Responsibilities:
  - Researched new methods to increase the performance of automatic volume control in virtual voice assistants
  - Implemented proposed method on large-scale computing resources from Amazon Web Services
- Project: Automatic volume control in virtual voice assistants
- Contributions:
  - Designed a new architecture that combined a neural network architecture and a gradient boosting on decision trees.
  - Utilized an interpretable method to understand the decision of the model
- Achievements:
  - The automatic volume control featured is launched in Amazon Alexa in 2021.
- Techniques and methodologies: Simple Recurrent Unit (SRU), LSTM, FFN, Catboost, SHapley Additive exPlanations (SHAP), log-filterbank energy, Adam optimization.
- Tools: Python, Pytorch

2016 - 2021 **City University of New York**, Research Assistant, US

- Start date: Aug/25/2016. End date: Dec/02/2021
- Advisor: Professor Michael Mandel
- Fundings and grants from 3 different organizations inside CUNY: CUNY Graduate Center, Brooklyn College, Research Foundation CUNY
- Responsibilities:
  - Researched new methodologies and techniques. Research topics: Speech recognition, speech enhancement and speech synthesis, acoustic modeling, audio signal processing, natural language processing, generative models, invariant representation, interpretable machine learning and deep learning.
  - Implemented and applied the proposed methods.
  - Published papers in conferences and journals.
  - Co-operated with other researchers.
- Projects:
  - ImportantAug: A data augmentation agent for speech

    * Introduced ImportantAug, a technique to augment training data for speech models by adding noise to unimportant regions of the speech and not to important regions.
    * On the standard Google Speech Commands test set, the proposed model achieved a 23.3% and 25.4% relative error rate reduction compared to conventional noise augmentation and a baseline without data augmentation, respectively.
    * ImportantAug significantly outperformed the conventional noise augmentation and the baseline on two test sets with additional noise (in-domain or out-of-domain noise) added.
    * Tools: Pytorch, Python, Torchaudio.
  - Direct comparison of the listening strategies in noise between humans and machines
    * Identified and compared the important time-frequency regions of human listeners and the automatic speech recognition (ASR).
    * Discovered that the important time-frequency regions of the time-delay neural network - long short-term memory networks acoustic model (TDNN-LSTM AM) were more similar to those of humans than the traditional Gaussian mixture model's regions (GMM).
    * The Jaccard similarity between human and GMM AM importance maps was 4.4% while human and neural network AM was 8.9%.
    * Recommended that the performance of ASR in noisy conditions could be improved by adapting it to attend to the same regions that humans use.
    * Tools: Matlab, Python, Kaldi. Dataset: AMI, CHIME-2 track 1.
  - Bubble cooperative networks for identifying important speech cues
    * Introduced a system called the Bubble cooperative network (BCN) consisting of a generator (LSTM) and a discriminator (LSTM) to identify important time-frequency regions of speech.
    * The BCN could obscure 97.7% the spectrogram with noise while maintaining recognition accuracy for a speech recognizer comparing a noisy test with a clean reference utterance.
    * The masks predicted by BCN showed patterns similar to analyses derived from human listening tests with better generalization and less context-dependence than other approaches.
    * Tools: Tensorflow, Python.
  - Concatenative analysis-by-synthesis
    * Utilized pitch and intensity information to improve the performance of a feed-forward neural network unit-selection in a concatenative speech synthesizer system, which is aimed at producing a high-quality clean speech from noisy speech.
  - Multi-channel speech enhancement
    * Deployed a baseline method, which estimated the noise covariance matrix for the beamforming to improve the far-field speech recognition.
    * Tools: Tensorflow, Python.

2011 - 2016 **Texas Instruments(TI)**, Technical Business Development Engineer, Vietnam
Start date: May/09/2011. End date: Aug/09/2016.

- Responsibilities:
  - Provided TI solutions and integrated circuit products for clients to build electronic devices: smart phone, telecom base station, set top box, smart home devices and toy robots.
  - Managed TI North Vietnam business development.
  - Conducted bi-weekly review with distributors: Avnet, Arrow, SS, WT and WPI.
- Achievements:
  - Increased TI North Vietnam revenue by 250% in 2012, 27% in 2013, 69% in 2014, 150% in 2015 and 30% in 2016.
  - Received a reward letter from TI Asia President for achievements in 2016.

Technical skills

PyTorch, Tensorflow, Keras, ESPnet, Kaldi, Horovod, Moses, Message Passing Interface.
Python, Matlab, C, C++, PHP, Java, Visual Basic, R, MySQL, HTML.

Service

Reviewer: Association for the Advancement of Artificial Intelligence (AAAI) 2021.

Subreviewer: International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018 and 2020, International Conference on Learning Representations (ICLR) 2019, International Speech Communication Association (Interspeech) 2019, Neural Information Processing Systems (NeurIPS) 2018, AAAI 2018.