Trung X. Pham
(Curriculum vitae)

CONTACT INFORMATION	**Full Name**: Pham Xuan Trung **Mobile**: (+84) 034-972-7598, (+82) 010-4390-8980 **Email**: phamxuantrungbk@gmail.com **Skype:** trung.pxt, **Kakao Talk**: trungpxvn **Homepage**: https://trungpx.github.io/	**Campus:** Korea Advanced Institute of Science and Technology (KAIST) **Address**: 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea **Major**: Electrical Engineering **School Email**: trungpx@kaist.ac.kr

GOOGLE SCHOLAR	**https://scholar.google.com/citations?user=4DkPIIAAAAAJ**
RESEARCH INTEREST	Artificial Intelligence/Machine Learning/Deep Learning, Computer Vision, Statistic and Analysis, Natural Language Processing, Speech Processing, Data Visualization.

EDUCATION	**Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea** **2018 – Now** School of Electrical Engineering **Ms-PhD** Electrical Engineering, studying on *Artificial Intelligent* (**ongoing**)
	Hanoi University of Science and technology (HUST), Hanoi, Vietnam **2009 – 2014** School of Electronics and Telecommunications (SET) **BS** Electronics and Communication Engineering, *June, 2014* • **GPA**: **3.52/4.00 (9.04 in scale 10)** • **Rank**: **10/526** students of department SET (**top 1.9%**) • **Rank**: **1/43** students in the undergraduate class
	Giao Thuy B high school, Nam Dinh, Viet Nam **2006 – 2009**

HONORS AND AWARDS	• **Award of dominant BS degree** of **HUST**, August, 2014 • **Annual Encourage Scholarship** by Ministry of Education and Training for excellent students with outstanding performance for every semester, 2009 – 2014 • **Trang Nguyen Flower** Award for the best student of Giao Thuy B high school, December, 2009 • **First award** in the Province Mathematic Contest for grade 12 students, September, 2009

WORKING AND RESEARCH EXPERIENCES	**1) Mobile engineer at VNPT Technology (in Hanoi)** **2014 - 2018** • Research and deploy 2G, 3G & 4G mobile communication network projects in a well-known telecom firm in Vietnam (VNPT Technology) with the various vendors such as Alcatel-Lucent, Nokia Siemens, SAMSUNG, PM and support for technical services in GSM mobile network. **2) Artificial Intelligent research at KAIST (Top University in South Korea)** **2018 – Now** **Working on projects related to:** Artificial Intelligence/Machine Learning/Deep Learning, Computer

Vision, Statistic and Analysis, Natural Language Processing, Speech Processing, Data Visualization.

3) Highlights
- **Have published in top-tier AI conferences and Journals:** ICML (1), ICLR (2), CVPR (3), ECCV (1), NeurIPS (1), Advanced Materials (1), Nano Energy (1), IEEE Access (3), TCSVT (1), TBC (1), ICASSP (1).
- **A research is selected as Top 100 best Korea national researches** in Korean casebook 2023 with **our research on AI & Acoustic Sensors** (certificated).
- **Reviewers and Program Committee for top-tier AI conferences and Journals:**
  - **ICML** (2024, 2025), **ICLR** (2024, 2025), **ICCV** 2023, **CVPR** (2023, 2024, 2025), **NeurIPS** (2023, 2024), **ICASSP** (2024, 2025), **ACCV** 2024, **AAAI** (2024, 2025), **ECCV** 2024, **AISTATS** (2025), **IJCNN** (2025).
  - **Neural Networks** (2023), **TMM** (2023), **CVIU** (2024), **ISPRS** (2024), **ESWA** (2024, 2025), **DSP** (2025), **EAAI** (2025), **TMLR** (2025).

Projects and Researches

1. **Speaker Recognition (AI and Flexible Acoustic Sensors)**
   - Input: Speech signal (waveform)
   - Output: Who is speaking? (predicted class) amongst total of 40 classes.
   - Method: End-to-end deep learning algorithm (CNN), using MFCC feature.
   - Performance: Reaching 99.3% test accuracy.
   - **Framework: Keras/tensorflow.**
   - Refer to publication paper below (**top journals**: **Advanced Materials and Nano Energy**)

2. **Benign Paroxysmal Positional Vertigo (BPPV) disorders diagnosis**
   - Input: Video (recorded body scene and movement of two eyes), the environment setting is conducted in the medical field.
   - Output: Types of disorders (6 classes)
   - Method: Deep learning algorithm (CNN, LSTM, attention models). Extracting feature from RGB frames by CNN for spatial stream, computing optical flow field for motion stream. Combine two of modalities into a final rich representation of the given video input for BPPV types recognition.
   - **Framework: Pytorch**
   - Refer to publication paper below (**IEEE Access journal**)

3. **Generative Adversarial Networks (GANs)**
   - Using Deep Convolutional Neural Network (GAN framework) to generate image by learning the distribution from an image dataset.
   - Now, AI model can create an entirely new realistic image of a person that does not exist in the

world.
- **Framework: Pytorch**

4. **Visual Question Answering (VQA)**
   - Using language model and deep learning to learn visual features of image and answer the questions. Input are image and question (text domain).
   - We proposed an attention mechanism so that it can focus to put more weights adaptively to either visual or language modality, so that can predict the answer accurately.
   - **Framework: Pytorch**

5. **Music Generation - Learning to generate high performance music by using RNN**
   - Directly learn the audio waveform by using RNN model to generate new piano music audio. We generated an output example by deep learning model as shown in the links below, computer performed as a real pianist: https://clyp.it/qjjz1kib, or https://clyp.it/yngyojt0
   - **Framework: Pytorch/Keras/Tensorflow**

6. **Video Question Answering (Video QA)**
   - Input: Video (from some real programs: medical dramas, sitcoms, and crime shows) + subtitle + question (text domain).
   - Output: Answer (text domain).
   - Method: CNN network for video stream, GloVE/BERT for text domain. Learn features representation by LSTM and attention mechanism.
   - **Framework: Pytorch**
   - Refer to publication paper below (**CVPR, top best conferences on AI**)

7. **Object Detection**
   - Using framework of Faster R-CNN (two stages detector), or YOLO (1 stage detector) for objects detection in images/video. Those are the state-of-the-art methods in computer vision for detecting 80 objects (COCO dataset) in the given image (cat, dog, person, car, airplane, …).
   - **Framework: Pytorch**
   - Refer to publication paper below (**NeurIPS, top best conferences on AI**)

8. **Self-Supervised Learning (SSL)**
   - In this work, the visual representation of the image has been learned completely in an unsupervised manner (without any manual label provided by human).
   - To this end, a pretext task that is designed by auto generated label for image by some image transforms in the local region of image. The network aims to predict which transformation has been applied into the image.
   - By solving the proposed proxy task, the CNN model (ResNet) can learn the good representation of image which can be used for the downstream tasks effectively.

	• **Framework: Pytorch** • Refer to published papers below (**ICLR, CVPR, top best conferences on AI**) **9. Generative AI with Diffusion Models** • In this work, we build a cutting-edge technique that can control the person image synthesis via conditional diffusion model. The proposed method shows the remarkable performance while being much more efficient compared to the existing techniques. • **Framework: Pytorch** • Please refer to the publication below (**ICML, ICLR, top best conferences on machine learning and AI**)
PUBLICATIONS	1. **Trung X. Pham**, Zhang Kang, Hong Ji Woo, Xuran Zheng, Chang D. Yoo, "E-MD3C: Taming Masked Diffusion Transformers for Efficient Zero-Shot Object Customization", **Under review** (**Arxiv 2025**). 2. **Trung X. Pham***, Tri Ton*, Chang D. Yoo, "MDSGen: Fast and Efficient Masked Diffusion Temporal-Aware Transformers for Open-Domain Sound Generation", 13rd International Conference on Learning Representations (**ICLR 2025**). 3. **Trung X. Pham***, Zhang Kang*, Chang D. Yoo, "Cross-view Masked Diffusion Transformers for Person Image Synthesis", Forty-first International Conference on Machine Learning (**ICML 2024**). 4. Axi Niu, **Trung X. Pham**, Zhang Kang, J Sun, Y Zhu, In So Kweon, Y Zhang, "ACDMSR: Accelerated Conditional Diffusion Models for Single Image Super-Resolution", IEEE Transactions on Broadcasting (**TBC 2024**) (**Q1, SCIE, IF: 5.194**). 5. Axi Niu, Zhang Kang, **Trung X. Pham**, P Wang, J Sun, In So Kweon, Y Zhang, "Learning from Multi-Perception Features for Real-Word Image Super-resolution", IEEE Transactions on Circuits and Systems for Video Technology (**TCSVT 2024**) (**Q1, SCIE, IF: 8.4**). 6. **Trung X. Pham**, Axi Niu, Zhang Kang, Ji Woo Hong, Joshua Tian Jin Tee, Chang D. Yoo, "Self-Supervised Visual Representation Learning via Residual Momentum", (**IEEE Access 2023**) (**Q1, SCIE, IF: 3.9**). 7. Axi Niu, Zhang Kang, **Trung X. Pham**, J Sun, Y Zhu, In So Kweon, Y Zhang, "cDPMSR: Conditional Diffusion Probabilistic Models for Single Image Super-Resolution", (**ICIP 2023**). 8. Thanh Nguyen***,** **Trung X. Pham*,** Chaoning Zhang, Tung Luu, Thang Vu, and Chang D. Yoo, "DimCL: Dimensional Contrastive Learning for Improving Self-Supervised Learning", (**IEEE Access 2023**) (**Q1, SCIE, IF: 3.9**). (***equal contribution**). 9. **Trung X. Pham,** Jin Woong Choi, Rusty Lloy Mina, Thanh Nguyen, Sultan Rizky Madjid, Chang D. Yoo, "LAD: A Deep Hybrid Learning-based System for Benign Paroxysmal Positional Vertigo Diseases Diagnostic", (**IEEE Access 2022**) (**Q1, SCIE, IF: 3.9**). 10. **Trung X. Pham**, Chaoning Zhang, Axi Niu, Kang Zhang, Chang D. Yoo, "On the Pros and Cons of Momentum Encoder in Self-Supervised Visual Representation Learning", **Arxiv 2022**. 11. Young Hoon Jung*, **Trung X. Pham***, Dias Issa, Hee Seung Wang, Jae Hee Lee, Mingi Chung, Bo-Yeon Lee, Gwangsu Kim, Chang D. Yoo, Keon Jae Lee, "Deep Learning-based Noise-Robust Flexible Piezoelectric Acoustic Sensors for Speech Processing", (**Nano Energy 2022**) (**Q1, SCIE, IF:**

19.07). (***equal contribution**).

12. Chaoning Zhang*, Kang Zhang*, **Trung X. Pham***, Axi Niu, Zhinan Qiao, Chang D. Yoo, In So Kweon, "Towards Understanding and Simplifying MoCo: Dual Temperature Helps Contrastive Learning without Many Negative Samples", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2022**) (***equal contribution**).

13. Chaoning Zhang, Kang Zhang, Chenshuang Zhang, **Trung X. Pham**, Chang D. Yoo, In So Kweon, "How Does SimSiam Avoid Collapse Without Negative Samples? A Unified Understanding with Contrastive Learning", The 10th International Conference on Learning Representations (**ICLR 2022**).

14. **Trung X. Pham**, Rusty John Lloyd Mina, Dias and Chang D. Yoo, "Self-supervised Learning with Local Attention-Aware Feature", **Arxiv 2021.**

15. Thanh Nguyen, Tung Luu, **Trung X. Pham**, Sanzhar Rakhimkul, Chang D. Yoo, "Robust MAML: Prioritization Task Buffer with Adaptive Learning Process for Model-Agnostic Meta-Learning", The International Conference on Acoustics, Speech, & Signal Processing (**ICASSP 2021**).

16. Donghoon Lee, Hyunsin Park, **Trung X. Pham**, Chang D. Yoo, "Learning Augmentation Network via Influence Functions", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2020**).

17. Junyeong Kim, Minuk Ma, **Trung X. Pham**, Kyungsu Kim, Chang D. Yoo, "Modality Shifting Attention Network for Multi-Modal Video Question Answering", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2020**).

18. Young Hoon Jung, Seong Kwang Hong, Hee Seung Wang, Jae Hyun Han, **Trung X. Pham**, Hyunsin P., Junyeong Kim, Sunghun K., Chang D. Yoo, Keon Jae Lee, "Flexible piezoelectric acoustic sensors and machine learning for speech processing", (**Advanced Materials 2020) (Q1, SCIE, IF: 32.09)**.

19. Thang Vu, Hyunjun Jang, **Trung X. Pham**, Chang D. Yoo, "Cascade RPN: Delving into High-Quality Region Proposal Network with Adaptive Convolution", Advances in Neural Information Processing Systems (**NeurIPS 2019**).

20. **Trung X. Pham**, Chang D. Yoo, "Short Convolutional Neural Network and MFCCs for Accurate Speakers Recognition Systems", The 34th International Technical Conference on Circuits/Systems, Computers and Communications, (**ITC-CSCC 2019**).

21. Thang Vu, Cao V Nguyen, **Trung X. Pham**, Tung M Luu, Chang D Yoo, "Fast and Efficient Image Quality Enhancement Via Desubpixel Convolutional Neural Networks", Proceedings of the European Conference on Computer Vision (**ECCV 2018**).

22. **Trung X. Pham**, Dias Issa, Chang D. Yoo, "The Era of Representation Learning without any Expensive Manual Labels via Contrastive Learning: A survey", 대한전자공학회 학술대회, (2021) p1146-1148.

23. Dias Issa, **Trung X. Pham**, Chang D. Yoo, "Self-Supervised Learning through Ad Hoc Pretext Tasks: Survey", 대한전자공학회 학술대회, (2021) p1153-1155.

24. **Trung X. Pham**, Chang D. Yoo, "Benign Paroxysmal Positional Vertigo Diseases Diagnostic by Using Artificial Intelligent", 2019 한국인공지능학회 추계학술대회.

PROJECTS	1. **Gene Sequence to detect 6 neurological diseases, 2018-2019**
	  • Diagnosis of six diseases based on the gene data
	2. **Medical Project, 2019-2021**
	  • Diagnosis of Benign Paroxysmal Positional Vertigo Disorders
	3. **Human Plus, 2018-2023**
	  • Developing machine learning/deep learning algorithms for new sensors with highly-sensitive signals for speech processing/speaker recognition. **Our research has been selected in casebook for top 100 best national researches of Korea in 2023 (certificated).**
	4. **AI 분야 SW StarLab, 2022-2023**
	  • Causality via Video Understanding in Machine Learning and Deep Learning

COMPUTER SKILLS	1. **Programming**: C/C++/C#, Python (Tensorflow/Pytorch/Keras/), MATLAB 2. **Operation Systems**: Windows: XP/7/8/10/11, Linux: Ubuntu, Unix: Sun Solaris 3. **Editors:** MS Office (PowerPoint, Excel, Word, Outlook, Visio), Latex.
LANGUAGES	1. **Vietnamese**: Native or bilingual proficiency 2. **English**: Full professional proficiency 3. **Korean**: Limited working proficiency
ACTIVITIES	1. Traveled for living, working, business, study, and tourism in 10 countries on Earth including Asia and America continental. 2. Tennis, Table Tennis, Badminton, GYM, Jogging/Hiking, Go Fishing

REFERENCES	**Professor** *Chang D. Yoo* • School of Electrical Engineering • Korea Advanced Institute of Science and Technology (KAIST) • Email: cd_yoo@kaist.ac.kr • Website: https://slsp.kaist.ac.kr/xe/, http://sanctusfactory.com/u-aim/ * Professor Chang D. Yoo is my advisor for Master & PhD program, at KAIST, Daejeon, S. Korea, **2018-2025**.
	Professor *Nguyen Hoang Hai* • Vice Dean Department of Telecommunication Systems • Hanoi University of Science and Technology (HUST) • Email: nhhaijp@mail.hut.edu.vn or nhhaijp@gmail.com • Phone: (+84) 905-555-399 • Website: https://set.hust.edu.vn/nguyen-hoang-hai * Prof. Nguyen Hoang Hai was my research and undergraduate advisor, at HUST, Hanoi, Vietnam, **2013-2014**.