HAI XUAN PHAM

Phone number: +44 7990 856337   27 Stourbridge Grove
pham.xuan.hai@outlook.com   Cambridge CB1 3HZ, UK
Website: https://haixpham.github.io

I am an experienced scientist with research interests lying at the intersection of machine learning, computer vision and natural language processing, aiming to solve a series of fundamental problems: develop models to learn structured multimodal representations of concepts, and make decisions through reasoning upon the learned knowledge. During the Ph.D. study, my research focused on multimodal human face and facial expression modelling, combining techniques from machine learning, computer vision, graphics and speech processing.

EDUCATION

Ph.D. Rutgers University, Department of Computer Science   Oct 2018
Dissertation: "Learning Human Facial Performance: Analysis and Synthesis"

B.Eng. Hanoi University of Science and Technology   Jul 2008
Information Technology, First Class

WORK EXPERIENCE

Samsung AI Center, Cambridge, UK   2019 - present
Research Scientist

- Research on multimodal (vision-language) understanding, and structured representation.
- Main research topics:
    - Cross-modal image-text retrieval and synthesis
    - Action flow graph extraction from procedural text
    - Action flow graph - video alignment
    - Graph-based procedural question-answering: In this work a series of LLMs were trained to generate text from graph representations, and to generate answers. The 3B-parameter model trained on our graph-based synthetic data far surpassing ChatGPT in question-answering performance.
    - Diffusion models: high-resolution image generation, textual inversion image editing.
    - Generative multimodal Model: develop a universal multimodal model that can take and generate data of different modalities.

Rutgers, The State University of New Jersey, USA   2011 to 2018
Department of Computer Science
Research Assistant

- Focused on multimodal face and continuous facial expression modeling, combining machine learning with computer vision, graphics and speech processing.
- Main research projects:
    - Real-time 3D blendshape-based face and facial expression tracking. A learned 3D face alignment model combined with inference time

optimization resulted in a robust 3D face tracker, capable of reliably tracking at varying distance. Worked with RGB and RGBD data.

- o Depth refinement with 3D face prior: jointly optimize 3D face model fitting and depth refinement to improve the quality of 3D reconstruction on noisy Kinect depth data.
- o 3D face animation from speech: learned a deep model for real-time 3D facial expression synthesis from speech audio.
- o Expressive face synthesis from still portraits: trained a GAN to synthesize novel expressions from any still photo of arbitrary person, capable of generating continuous articulated facial animation.

Agribank, Hanoi, Vietnam   2009 to 2011
Software Engineer
- Developed the core banking system
- Used Pro*C and Oracle for high performance data processing and analysis.

Toshiba Software Development, Hanoi, Vietnam   2008 to 2009
Software Engineer
- Developed embedded systems for Toshiba devices.

PUBLICATIONS

H. Yang, A. Bulat, I. Hadji, H. X. Pham, X. Zhu, G. Tzimiropoulos and B. Martinez, "FAM Diffusion: Frequency and Attention Modulation for High-Resolution Image Generation with Stable Diffusion", to appear in CVPR 2025.

H. X. Pham, I. Hadji, X. Xu, Z. Degutyte, J. Rainey, E. Kazakos, A. Fazly, G. Tzimiropoulos and B. Martinez, "Graph Guided Question Answer Generation for Procedural Question-Answering", in EACL 2024 (oral presentation).

N. Dvornik, I. Hadji, H. X. Pham, D. Bhatt, B. Martinez, A. Fazly and Allan D. Jepson, "Flow Graph to Video Grounding for Weakly-Supervised Multi-step Localization", in ECCV 2022 (oral presentation).

M. Kim, R. Guerrero, H. X. Pham and V. Pavlovic, "Variational Continual Proxy-Anchor for Deep Metric Learning", in International Conference on Artificial Intelligence and Statistics (AISTATS) 2022.

R. Guerrero, H. X. Pham and V. Pavlovic, "Cross-modal Retrieval and Synthesis (X-MRS): Closing the Modality Gap in Shared Subspace Learning", in ACM Multimedia 2021.

H. X. Pham, R. Guerrero, J. Li and V. Pavlovic, "CHEF: Cross-modal Hierarchical Embeddings for Food Domain Retrieval", in AAAI 2021.

H. X. Pham, Y. Wang and V. Pavlovic, "Learning Continuous Facial Actions From Speech for Real-Time Animation", in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1567-1580, 1 July-Sept. 2022 (accepted for publication in 2020).

H. X. Pham. Y. Wang and V. Pavlovic, **"**Generative Adversarial Talking Head: Bringing Portrait To Life with a Weakly Supervised Neural Network", in arXiv, 2018.

H. X. Pham, Y. Wang and V. Pavlovic, "End-to-end Learning for 3D Facial Animation from Speech", in International Conference in Multimodal Interaction (ICMI), Oct 2018.

C. Chen, H. X. Pham, V. Pavlovic, J. Cai, G. Shi, Y. Gao and H. Cheng, "Using 3D Face Priors for Depth Recovery", in Journal of Visual Communication and Image Representation, Vol 48, Oct 2017.

H. X. Pham, S. Cheung and V. Pavlovic, "Speech-driven 3D Facial Animation with Implicit Emotional Awareness: A Deep Learning Approach", in CVPRW 2017.

H. X. Pham and V. Pavlovic, "Robust Real-Time 3D Face Tracking from RGBD Videos under Extreme Pose, Depth, and Expression Variations", in International Conference in 3D Vision (3DV), 2016.

H. X. Pham, V. Pavlovic, J. Cai and T. Cham, "Robust Real-time Performance-driven 3D Face Tracking", in ICPR, 2016.

C. Chen, H. X. Pham, V. Pavlovic, J. Cai and G. Shi, "Depth Recover with Face Priors", in ACCV, 2014 (oral presentation).

H. X. Pham and V. Pavlovic, "Hybrid On-line 3D Face and Facial Actions Tracking in RGBD Video Sequences", in ICPR, 2014.

PATENTS

Hai Xuan Pham, Isma Hadji, Xinnuo Xu, Ziedune Degutyte, Jay Rainey, Evangelos Kazakos, Afsaneh Fazly, Georgios Tzimiropoulos, Brais Martinez Alonso, "METHOD FOR QUESTION-ANSWER GENERATION", UK Patent Publication GB2631848.

Vladimir Pavlovic, Minyoung Kim, Ricardo Guerrero, Hai Xuan Pham, "METHOD AND APPARATUS FOR CONCEPT MATCHING", US Patent US-20230137671-A1.

TEACHING EXPERIENCE

Rutgers, The State University of New Jersey, USA
Department of Computer Science

Teaching Assistant   2013 - 2018
- Software Methodology
- System Programming
- Introduction to Discrete Structures I (Set Theory and Logics)
- Introduction to Discrete Structures II (Probability Theory)
- Compilers

Lecturer   Summer 2017
- Introduction to Discrete Structures II (Probability Theory)

SCHOLARSHIPS AND AWARDS

Vietnam Education Fellowship   2011
Granted for 2-year of graduate study at Rutgers University (2011-2013)

Rutgers Outstanding Graduate Fellowship 2017
Granted allowance & tuition fee from Rutgers Graduate School.

Nvidia GPU Research Grant 2017
Granted GPUs from Nvidia to conduct research on multimodal face analysis.

PROFESSIONAL SERVICE

Peer-reviewed for: CVPR, ICCV, ECCV, ACCV, NeurIPS, ICLR, ACCV, BMVC, WACV, AISTATS, KDD, AAAI, TPAMI, TAC, etc.

TECHNICAL SKILLS

Programming: C/C++ (15+ year experience – circa 2018), Python (18+ year experience), Matlab, C#, Java, etc.

Platforms: Windows, Linux, Android

Deep learning: Pytorch and associated frameworks: Huggingface Transformers/Accelerate/Peft (maintaining my own forks), MS Deepspeed, etc.. I also have experience with scaling multi-node training. During my Ph.D., I mostly used the now discontinued CNTK deep learning toolkit in my research.

OTHER

Citizenship: Vietnamese
Residence: UK Indefinite Leave to Remain

REFERENCES

Provided upon request.