MOVING FROM FUNDAMENTALS OF COMPUTER VISION TO MEDICAL AI APPLICATIONS

A PERSONNEL NOTE

Hieu H. Pham, Ph.D.
Assistant Professor, College of Engineering & Computer Science,
VinUni-Illinois Smart Health Center, VinUniversity,
Visiting Scholar, Coordinated Science Laboratory, University of Illinois Urbana-Champaign,
E-mail: hieu.ph@vinuni.edu.vn
Homepage: https://huyhieupham.github.io/

April 21, 2023

ABSTRACT

In the summer of 2015, I packed my bags and left my hometown of Hanoi, Vietnam, to pursue a research internship in Grenoble, France. I began learning the fundamentals of computer vision and was fascinated by the potential applications of this technology. Over the next few years, I focused on human action recognition and behavior understanding. I worked on developing algorithms that could analyze video footage and identify patterns in human behavior. It was an exciting time to be working in this field, as advances in machine learning were making it possible to analyze and understand human behavior in ways that were previously impossible. In 2019, I started to explore the use of AI for medical applications. I began attending medical conferences and networking with researchers in the field. I was struck by the potential of using machine learning to diagnose and treat diseases. I started working on developing algorithms that could analyze medical images and detect early signs of diseases. It was a challenging transition, but I was determined to succeed. I spent long hours studying medical imaging data and learning about the various techniques and tools used in medical research. I collaborated with medical professionals to gain insights into the clinical context of my work. Finally, my hard work paid off. I was offered a position in a medical research lab where I could apply my skills to the development of AI algorithms for medical applications. I was thrilled to be working on a project that could have a real impact on people's lives. Today, I continue to work on the development of AI algorithms for medical applications. I am grateful for the journey that brought me from Hanoi to Grenoble, and for the opportunities that I have had to make a difference through my research. I look forward to the possibilities that lie ahead as I continue to explore the intersection of AI and medicine.

1 Introduction

I am currently an Assistant Professor at the College of Engineering and Computer Science (CECS), VinUniversity, and a Research Fellow cum Associate Director at VinUni-Illinois Smart Health Center. I received his Ph.D. in Computer Science from the Toulouse Computer Science Research Institute (IRIT), University of Toulouse, France, in 2019. Previously, I earned the Degree of Engineer in Industrial Informatics from Hanoi University of Science and Technology (HUST), Vietnam, in 2016. My research interests include Computer Vision, Machine Learning, Medical Image Analysis, and their applications in Smart Healthcare. Before joining VinUniversity, I worked at Vingroup Big Data Institute (VinBigData) as a Research Scientist and Head of the Fundamental Research Team. With this position, I led several research projects on Medical AI, including collecting various types of medical data, managing and annotating data, and developing new AI solutions for medical analysis.

2 Research interests

My research interests include Artificial Intelligence (AI), Machine Learning, Deep Learning, Computer Vision, especially their applications in Smart Healthcare, e.g. Medical Imaging Diagnosis, AI-based Computer-aided Diagnosis (AI-CAD), AI-assisted Diagnosis and Treatment, AI-assisted Disease Prevention and Risk Monitoring.

3 My research journey

We have developed a software for the detection and characterisation of defects based on the analysis of 3D point clouds provided by a scanner. This software has been developed within an industrial application dealing with the control of an aircraft fuselage surface. We then published the paper titled "Detection and characterization of surface defects based on the analysis of 3D point clouds provided by a scanner" Jovancevic et al. [a,b] in 2016. Détection et caractérisation de défauts de surface par analyse des nuages de points 3D fournis par un scanner Jovančević et al. [2017a]. Détection et caractérisation de défauts par analyse des nuages de points 3D fournis par un scanner Jovančević et al. [2017b]. 3D point cloud analysis for detection and characterization of defects on airplane exterior surface Jovančević et al. [2017c].

3.1 Human Action Recognition and Behavior Understanding

My very first paper was on human fall detection using RGB-D camera. It was "Real-time obstacle detection system in indoor environment for the visually impaired using microsoft kinect sensor" Pham et al. [2016]. Any mobility aid for the visually impaired people should be able to accurately detect and warn about nearly obstacles. In this paper, we present a method for support system to detect obstacle in indoor environment based on Kinect sensor and 3D-image processing. Color-Depth data of the scene in front of the user is collected using the Kinect with the support of the standard framework for 3D sensing OpenNI and processed by PCL library to extract accurate 3D information of the obstacles. The experiments have been performed with the dataset in multiple indoor scenarios and in different lighting conditions. Results showed that our system is able to accurately detect the four types of obstacle: walls, doors, stairs, and a residual class that covers loose obstacles on the floor. Precisely, walls and loose obstacles on the floor are detected in practically all cases, whereas doors are detected in 90.69% out of 43 positive image samples. For the step detection, we have correctly detected the upstairs in 97.33% out of 75 positive images while the correct rate of downstairs detection is lower with 89.47% from 38 positive images. Our method further allows the computation of the distance between the user and the obstacles. In June 2018, my paper "Exploiting deep residual networks for human action recognition from skeletal data" Pham et al. [2018a] has been accepted for publication in the Computer Vision and Image Understanding Journal. The computer vision community is currently focusing on solving action recognition problems in real videos, which contain thousands of samples with many challenges. In this process, Deep Convolutional Neural Networks (D-CNNs) have played a significant role in advancing the state-of-the-art in various vision-based action recognition systems. Recently, the introduction of residual connections in conjunction with a more traditional CNN model in a single architecture called Residual Network (ResNet) has shown impressive performance and great potential for image recognition tasks. In this paper Pham et al. [2018a], we investigate and apply deep ResNets for human action recognition using skeletal data provided by depth sensors. Firstly, the 3D coordinates of the human body joints carried in skeleton sequences are transformed into image-based representations and stored as RGB images. These color images are able to capture the spatial-temporal evolutions of 3D motions from skeleton sequences and can be efficiently learned by D-CNNs. We then propose a novel deep learning architecture based on ResNets to learn features from obtained color-based representations and classify them into action classes. The proposed method is evaluated on three challenging benchmark datasets including MSR Action 3D, KARD, and NTU-RGB+D datasets. Experimental results demonstrate that our method achieves state-of-the-art performance for all these benchmarks whilst requiring less computation resource. In particular, the proposed method surpasses previous approaches by a significant margin of 3.4% on MSR Action 3D dataset, 0.67% on KARD dataset, and 2.5% on NTU-RGB+D dataset. The paper now reached more than 80 citations (April, 2023). I think this work is the biggest one during my Ph.D. life (2016–2019).

Automatic human action recognition is indispensable for almost artificial intelligent systems such as video surveillance, human-computer interfaces, video retrieval, etc. Despite a lot of progress, recognizing actions in an unknown video is still a challenging task in computer vision. Recently, deep learning algorithms have proved its great potential in many vision-related recognition tasks. In this paper, we propose the use of Deep Residual Neural Networks (ResNets) to learn and recognize human action from skeleton data provided by Kinect sensor. Firstly, the body joint coordinates are transformed into 3D-arrays and saved in RGB images space. Five different deep learning models based on ResNet have been designed to extract image features and classify them into classes. Experiments are conducted on two public video datasets for human action recognition containing various challenges. The results show that our method achieves the state-of-the-art performance comparing with existing approaches. We then propose the use of Deep Residual Neural Networks (ResNets) to learn and recognize human action from skeleton data provided by Kinect sensor Pham [2017].

Firstly, the body joint coordinates are transformed into 3D-arrays and saved in RGB images space. Five different deep learning models based on ResNet have been designed to extract image features and classify them into classes. Experiments are conducted on two public video datasets for human action recognition containing various challenges. The results show that our method achieves the state-of-the-art performance comparing with existing approaches. We propose a novel skeleton-based representation for 3D action recognition in videos using Deep Convolutional Neural Networks (D-CNNs). Two key issues have been addressed: First, how to construct a robust representation that easily captures the spatial-temporal evolutions of motions from skeleton sequences. Second, how to design D-CNNs capable of learning discriminative features from the new representation in a effective manner. To address these tasks, a skeleton-based representation, namely, SPMF (Skeleton Pose-Motion Feature) is proposed. The SPMFs are built from two of the most important properties of a human action: postures and their motions. Therefore, they are able to effectively represent complex actions. For learning and recognition tasks, we design and optimize new D-CNNs based on the idea of Inception Residual networks to predict actions from SPMFs. Our method is evaluated on two challenging datasets including MSR Action3D and NTU-RGB+D. Experimental results indicated that the proposed method surpasses state-of-the-art methods whilst requiring less computation. Skeletal movement to color map: A novel representation for 3D action recognition with inception residual networks Pham et al. [2018b]. We present a new deep learning approach for real-time 3D human action recognition from skeletal data and apply it to develop a vision-based intelligent surveillance system. Given a skeleton sequence, we propose to encode skeleton poses and their motions into a single RGB image. An Adaptive Histogram Equalization (AHE) algorithm is then applied on the color images to enhance their local patterns and generate more discriminative features. For learning and classification tasks, we design Deep Neural Networks based on the Densely Connected Convolutional Architecture (DenseNet) to extract features from enhanced-color images and classify them into classes. Experimental results on two challenging datasets show that the proposed method reaches state-of-the-art accuracy, whilst requiring low computational time for training and inference. This paper also introduces CEMEST, a new RGB-D dataset depicting passenger behaviors in public transport. It consists of 203 untrimmed real-world surveillance videos of realistic normal and anomalous events. We achieve promising results on real conditions of this dataset with the support of data augmentation and transfer learning techniques. This enables the construction of real-world applications based on deep learning for enhancing monitoring and security in public transport. A deep learning approach for real-time 3D human action recognition from skeletal data Pham et al. [2019a]. Recognizing human actions in untrimmed videos is an important challenging task. An effective 3D motion representation and a powerful learning model are two key factors influencing recognition performance. In this paper we introduce a new skeleton-based representation for 3D action recognition in videos. The key idea of the proposed representation is to transform 3D joint coordinates of the human body carried in skeleton sequences into RGB images via a color encoding process. By normalizing the 3D joint coordinates and dividing each skeleton frame into five parts, where the joints are concatenated according to the order of their physical connections, the color-coded representation is able to represent spatio-temporal evolutions of complex 3D motions, independently of the length of each sequence. We then design and train different Deep Convolutional Neural Networks (D-CNNs) based on the Residual Network architecture (ResNet) on the obtained image-based representations to learn 3D motion features and classify them into classes. Our method is evaluated on two widely used action recognition benchmarks: MSR Action3D and NTU-RGB+D, a very large-scale dataset for 3D human action recognition. The experimental results demonstrate that the proposed method outperforms previous state-of-the-art approaches whilst requiring less computation for training and prediction. Learning to recognise 3D human action from a new skeleton-based representation using deep convolutional neural networks Pham et al. [2019b]. Designing motion representations for 3D human action recognition from skeleton sequences is an important yet challenging task. An effective representation should be robust to noise, invariant to viewpoint changes and result in a good performance with low-computational demand. Two main challenges in this task include how to efficiently represent spatio-temporal patterns of skeletal movements and how to learn their discriminative features for classification tasks. This paper presents a novel skeleton-based representation and a deep learning framework for 3D action recognition using RGB-D sensors. We propose to build an action map called SPMF (Skeleton Posture-Motion Feature), which is a compact image representation built from skeleton poses and their motions. An Adaptive Histogram Equalization (AHE) algorithm is then applied on the SPMF to enhance their local patterns and form an enhanced action map, namely Enhanced-SPMF. For learning and classification tasks, we exploit Deep Convolutional Neural Networks based on the DenseNet architecture to learn directly an end-to-end mapping between input skeleton sequences and their action labels via the Enhanced-SPMFs. The proposed method is evaluated on four challenging benchmark datasets, including both individual actions, interactions, multiview and large-scale datasets. The experimental results demonstrate that the proposed method outperforms previous state-of-the-art approaches on all benchmark tasks, whilst requiring low computational time for training and inference. Spatio–temporal image representation of 3D skeletal movements for view-invariant action recognition with deep convolutional neural networks Pham et al. [2019c]. We propose a novel skeleton-based representation for 3D action recognition in videos using Deep Convolutional Neural Networks (D-CNNs). Two key issues have been addressed: First, how to construct a robust representation that easily captures the spatial-temporal evolutions of motions from skeleton sequences. Second, how to design D-CNNs capable of learning discriminative features from the new representation in a effective manner. To

address these tasks, a skeletonbased representation, namely, SPMF (Skeleton Pose-Motion Feature) is proposed. The SPMFs are built from two of the most important properties of a human action: postures and their motions. Therefore, they are able to effectively represent complex actions. For learning and recognition tasks, we design and optimize new D-CNNs based on the idea of Inception Residual networks to predict actions from SPMFs. Our method is evaluated on two challenging datasets including MSR Action3D and NTU-RGB+D. Experimental results indicated that the proposed method surpasses state-of-the-art methods whilst requiring less computation. Skeletal Movement to Enhanced Color Map: A Novel Representation for Rgb-d Based 3D Human Action Recognition with Densely Connected Convolutional Networks Pham et al. [2019d]. Human action recognition is an important application domain in computer vision. Its primary aim is to accurately describe human actions and their interactions from a previously unseen data sequence acquired by sensors. The ability to recognize, understand, and predict complex human actions enables the construction of many important applications such as intelligent surveillance systems, human-computer interfaces, health care, security, and military applications. In recent years, deep learning has been given particular attention by the computer vision community. This paper presents an overview of the current state-of-the-art in action recognition using video analysis with deep learning techniques. We present the most important deep learning models for recognizing human actions, and analyze them to provide the current progress of deep learning algorithms applied to solve human action recognition problems in realistic videos highlighting their advantages and disadvantages. Based on the quantitative analysis using recognition accuracies reported in the literature, our study identifies state-of-the-art deep architectures in action recognition and then provides current trends and open problems for future works in this field. Video-based human action recognition using deep learning: a review Pham et al. [2022a]. We present a deep learning-based multitask framework for joint 3D human pose estimation and action recognition from RGB video sequences. Our approach proceeds along two stages. In the first, we run a real-time 2D pose detector to determine the precise pixel location of important keypoints of the body. A two-stream neural network is then designed and trained to map detected 2D keypoints into 3D poses. In the second, we deploy the Efficient Neural Architecture Search (ENAS) algorithm to find an optimal network architecture that is used for modeling the spatio-temporal evolution of the estimated 3D poses via an image-based intermediate representation and performing action recognition. Experiments on Human3.6M, MSR Action3D and SBU Kinect Interaction datasets verify the effectiveness of the proposed method on the targeted tasks. Moreover, we show that our method requires a low computational budget for training and inference. A unified deep framework for joint 3d pose estimation and action recognition from a single rgb camera Pham et al. [2020]

My PhD these titled "Deep learning architectures for human action recognition from monocular RGB-D video sequences. Application to public transport monitoring." Pham et al. [2015] under four advisor Khoudour, Louahdi and Crouzil, Alain and Zegers, Pablo and Velastin, Sergio A. This thesis is dealing with automatic recognition of human actions from monocular RGB-D video sequences. Our main goal is to recognize which human actions occur in unknown videos. This problem is a challenging task due to a number of obstacles caused by the variability of the acquisition conditions, including the lighting, the position, the orientation and the field of view of the camera, as well as the variability of actions which can be performed differently, notably in terms of speed. To tackle these problems, we first review and evaluate the most prominent stateof-the-art techniques to identify the current state of human action recognition in videos. We then propose a new approach for skeleton-based action recognition using Deep Neural Networks (DNNs). Two key questions have been addressed. First, how to efficiently represent the spatio-temporal patterns of skeletal data for fully exploiting the capacity in learning high-level representations of Deep Convolutional Neural Networks (D-CNNs). Second, how to design a powerful D-CNN architecture that is able to learn discriminative features from the proposed representation for classification task. As a result, we introduce two new 3D motion representations called SPMF (Skeleton Posture-Motion Feature) and Enhanced-SPMF that encode skeleton poses and their motions into color images. For learning and classification tasks, we design and train different D-CNN architectures based on the Residual Network (ResNet), Inception-ResNet-v2, Densely Connected Convolutional Network (DenseNet) and Efficient Neural Architecture Search (ENAS) to extract robust features from color–coded images and classify them. Experimental results on various public and challenging human action recognition datasets (MSR Action3D, Kinect Activity Recognition Dataset, SBU Kinect Interaction, and NTU-RGB+D) show that the proposed approach outperforms current state-of-the-art. We also conducted research on the problem of 3D human pose estimation from monocular RGB video sequences and exploited the estimated 3D poses for recognition task. Specifically, a deep learning-based model called OpenPose is deployed to detect 2D human poses. A DNN is then proposed and trained for learning a 2D-to-3D mapping in order to map the detected 2D keypoints into 3D poses. Our experiments on the Human3.6M dataset verified the effectiveness of the proposed method. These obtained results allow opening a new research direction for human action recognition from 3D skeletal data, when the depth cameras are failing. In addition, we collect and introduce in this thesis, CEMEST database, a new RGB-D dataset depicting passengers' behaviors in public transport. It consists of 203 untrimmed real-world surveillance videos of realistic "normal" and "abnormal" events. We achieve promising results on CEMEST with the support of data augmentation and transfer learning techniques. This enables the construction of real-world applications based on deep learning for enhancing public transportation management services. The thesis is available at `https://huyhieupham.github.io/data/Huy_Hieu_PHAM_Doctoral_Thesis.pdf`.

3.2 Building Medical Imaging Datasets

VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography Pham et al. [a]

3.3 Medical AI Research

VinDr-SpineXR: A deep learning framework for spinal lesions detection and classification from radiographs Nguyen et al. [2021a]. Learning to automatically diagnose multiple diseases in pediatric chest radiographs using deep convolutional neural networks Tran et al. [2021]. Dicom imaging router: An open deep learning framework for classification of body parts from dicom x-ray scans **?**. VinDr-SpineXR: A large annotated medical image dataset for spinal lesions detection and classification from radiographs Pham et al. [b]. A clinical validation of VinDr-CXR, an AI system for detecting abnormal chest radiographs Nguyen et al. [2021b]. Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels Pham et al. [2021]. VinDr-RibCXR: A benchmark dataset for automatic segmentation and labeling of individual ribs on chest X-rays Nguyen et al. [2021a]. VinDr-PCXR: An open, large-scale chest radiograph dataset for interpretation of thoracic diseases in children Nguyen et al. [2022a]. VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography Nguyen et al. [2022b]. VinDr-PCXR: An open, large-scale pediatric chest X-ray dataset for interpretation of common thoracic diseases Nguyen. Deployment and validation of an AI system for detecting abnormal chest radiographs in clinical settings Nguyen et al. [2022c]. Transparency strategy-based data augmentation for BI-RADS classification of mammograms Tran et al. [2022]. Phase recognition in contrast-enhanced CT scans based on deep learning and random sampling **?**. A novel multi-view deep learning approach for BI-RADS and density assessment of mammograms Nguyen et al. [2022d].VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations Nguyen et al. [2022e]. Slice-level Detection of Intracranial Hemorrhage on CT Using Deep Descriptors of Adjacent Slices Ngo et al. [2022]. An Accurate and Explainable Deep Learning System Improves Interobserver Agreement in the Interpretation of Chest Radiograph Pham et al. [2022b]. Learning to diagnose common thorax diseases on chest radiographs from radiology reports in Vietnamese Nguyen et al. [2022f]. Detecting COVID-19 from digitized ECG printouts using 1D convolutional neural networks Nguyen et al. [2022g]. A novel deep learning-based approach for sleep apnea detection using single-lead ECG signals Nguyen et al. [2022d]. Enhancing Few-shot Image Classification with Cosine Transformer Nguyen et al. [2022h]. Learning from multiple expert annotators for enhancing anomaly detection in medical image analysis Le et al. [2023]. Ensemble Learning of Myocardial Displacements for Myocardial Infarction Detection in Echocardiography Tuan et al. [2023].

4 Federated Learning For Healthcare Applications

Backdoor Attacks and Defenses in Federated Learning: Survey, Challenges and Future Research Directions Dung Nguyen et al. [2023]. Personalized Privacy-Preserving Framework for Cross-Silo Federated Learning Tran et al. [2023]. CADIS: Handling Cluster-skewed Non-IID Data in Federated Learning with Clustered Aggregation and Knowledge DIStilled Regularization Nguyen et al. [2023a].

5 Multimodal Biomedical AI

The increasing availability of biomedical data from large biobanks, electronic health records, medical imaging, wearable and ambient biosensors, and the lower cost of genome and microbiome sequencing have set the stage for the development of multimodal artificial intelligence solutions that capture the complexity of human health and disease. We outline the key applications enabled, along with the technical and analytical challenges. We explore opportunities in personalized medicine, digital clinical trials, remote monitoring and care, pandemic surveillance, digital twin technology and virtual health assistants. Further, we survey the data, modeling and privacy challenges that must be overcome to realize the full potential of multimodal artificial intelligence in health. Most of the current applications of AI in medicine have addressed narrowly defined tasks using one data modality, such as a computed tomography (CT) scan or retinal photograph. In contrast, clinicians process data from multiple sources and modalities when diagnosing, making prognostic evaluations and deciding on treatment plans. Furthermore, current AI assessments are typically one-off snapshots, based on a moment of time when the assessment is performed, and therefore not 'seeing' health as a continuous state. In theory, however, AI models should be able to use all data sources typically available to clinicians, and even those unavailable to most of them (for example, most clinicians do not have a deep understanding of genomic medicine). The development of multimodal AI models that incorporate data across modalities—including biosensors, genetic, epigenetic, proteomic, microbiome, metabolomic, imaging, text, clinical, social determinants and environmental data—is poised to partially bridge this gap and enable broad applications that include individualized medicine, integrated, real-time pandemic surveillance, digital clinical trials and virtual health coaches (Fig. 1). In this

Review, we explore the opportunities for such multimodal datasets in healthcare; we then discuss the key challenges and promising strategies for overcoming these. Basic concepts in AI and machine learning will not be discussed here but are reviewed in detail elsewhere (Multimodal biomedical AI Paper). High Accurate and Explainable Multi-Pill Detection Framework with Graph Neural Network-Assisted Multimodal Data Fusion Nguyen et al. [2023b]. Multi-stream Fusion for Class Incremental Learning in Pill Image Classification Nguyen et al. [2022i]. FedDCT: Federated Learning of Large Convolutional Neural Networks on Resource Constrained Devices using Divide and Co-Training Nguyen et al. [2022j]. FedDRL: Deep Reinforcement Learning-based Adaptive Aggregation for Non-IID Data in Federated Learning Nguyen et al. [2022k]. Multi-stream Fusion for Class Incremental Learning in Pill Image Classification Nguyen et al. [2022i].

6 What's next?

7 Concluding remarks

References

Igor Jovancevic, Huy-Hieu Pham, Jean-José Orteu, Rémi Gilblas, Jacques Harvent, Xavier Maurice, and Ludovic Brèthes. Detection and characterization of surface defects based on the analysis of 3d point clouds provided by a scanner detection and characterization of surface defects based on the analysis of 3d point clouds provided by a scanner. a.

I Jovancevic, H-H Pham, J-J Orteu, R Gilblas, J Harvent, X Maurice, and L Brèthes. Detection et caracterisation de defauts par analyse des nuages de points 3d fournis par un scanner. b.

Igor Jovančević, Huy-Hieu Pham, Jean-José Orteu, Rémi Gilblas, Jacques Harvent, Xavier Maurice, and Ludovic Brèthes. Détection et caractérisation de défauts de surface par analyse des nuages de points 3d fournis par un scanner. *Instrumentation, Mesure, Métrologie*, 16(1-4):p–261, 2017a.

Igor Jovančević, H-H Pham, Jean-José Orteu, Rémi Gilblas, J Harvent, X Maurice, and L Brèthes. Détection et caractérisation de défauts par analyse des nuages de points 3d fournis par un scanner. In *15ème Colloque Méthodes et Techniques Optiques pour l'Industrie, Le Mans (France), 20-24 mars 2017.*, 2017b.

Igor Jovančević, Huy-Hieu Pham, Jean-José Orteu, Rémi Gilblas, Jacques Harvent, Xavier Maurice, and Ludovic Brèthes. 3d point cloud analysis for detection and characterization of defects on airplane exterior surface. *Journal of Nondestructive Evaluation*, 36:1–17, 2017c.

Huy-Hieu Pham, Le Thi-Lan, and Nicolas Vuillerme. Real-time obstacle detection system in indoor environment for the visually impaired using microsoft kinect sensor. *Journal of Sensors*, 2016, 2016.

Huy-Hieu Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. Exploiting deep residual networks for human action recognition from skeletal data. *Computer Vision and Image Understanding*, 170:51–66, 2018a.

Huy-Hieu Pham. Learning and recognizing human action from skeleton movement with deep residual neural networks. 2017.

Huy-Hieu Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. Skeletal movement to color map: A novel representation for 3d action recognition with inception residual networks. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3483–3487. IEEE, 2018b.

Huy Hieu Pham, Houssam Salmane, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. A deep learning approach for real-time 3d human action recognition from skeletal data. In *Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part I 16*, pages 18–32. Springer, 2019a.

Huy-Hieu Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. Learning to recognise 3d human action from a new skeleton-based representation using deep convolutional neural networks. *IET Computer Vision*, 13(3):319–328, 2019b.

Huy Hieu Pham, Houssam Salmane, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. Spatio–temporal image representation of 3d skeletal movements for view-invariant action recognition with deep convolutional neural networks. *Sensors*, 19(8):1932, 2019c.

Huy-Hieu Pham, Houssam Salmane, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. Skeletal movement to enhanced color map: A novel representation for rgb-d based 3d human action recognition with densely connected convolutional networks. In *16th International Conference on Image Analysis and Recognition (ICIAR 2019), august 27-29, 2019, Waterloo, Canada*, 2019d.

Hieu H Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. Video-based human action recognition using deep learning: a review. *arXiv preprint arXiv:2208.03775*, 2022a.

Huy Hieu Pham, Houssam Salmane, Louahdi Khoudour, Alain Crouzil, Sergio A Velastin, and Pablo Zegers. A unified deep framework for joint 3d pose estimation and action recognition from a single rgb camera. *Sensors*, 20(7):1825, 2020.

Huy-Hieu Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, and Sergio A Velastin. Video-based human action recognition using deep learning. 2015.

Hieu Huy Pham, Hieu Nguyen Trung, and Ha Quy Nguyen. Vindr-mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography. a.

Hieu T Nguyen, Hieu H Pham, Nghia T Nguyen, Ha Q Nguyen, Thang Q Huynh, Minh Dao, and Van Vu. Vindr-spinexr: A deep learning framework for spinal lesions detection and classification from radiographs. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 291–301. Springer, 2021a.

Thanh T Tran, Hieu H Pham, Thang V Nguyen, Tung T Le, Hieu T Nguyen, and Ha Q Nguyen. Learning to automatically diagnose multiple diseases in pediatric chest radiographs using deep convolutional neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3314–3323, 2021.

Hieu Huy Pham, Hieu Nguyen Trung, and Ha Quy Nguyen. Vindr-spinexr: A large annotated medical image dataset for spinal lesions detection and classification from radiographs. b.

Ngoc Huy Nguyen, Ha Quy Nguyen, Nghia Trung Nguyen, Thang Viet Nguyen, Hieu Huy Pham, and Tuan Ngoc-Minh Nguyen. A clinical validation of vindr-cxr, an ai system for detecting abnormal chest radiographs. *arXiv preprint arXiv:2104.02256*, 2021b.

Hieu H Pham, Tung T Le, Dat Q Tran, Dat T Ngo, and Ha Q Nguyen. Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing*, 437:186–194, 2021.

Huy Ngoc Nguyen, Hieu Pham, Thanh Tran, Tuan Nguyen, and Quy Ha Nguyen. Vindr-pcxr: An open, large-scale chest radiograph dataset for interpretation of thoracic diseases in children. *medRxiv*, pages 2022–03, 2022a.

Hieu Trung Nguyen, Ha Quy Nguyen, Hieu Huy Pham, Khanh Lam, Linh Tuan Le, Minh Dao, and Van Vu. Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *MedRxiv*, pages 2022–03, 2022b.

Ha Quy Nguyen. Vindr-pcxr: An open, large-scale pediatric chest x-ray dataset for interpretation of common thoracic diseases.

Ngoc Huy Nguyen, Ha Quy Nguyen, Nghia Trung Nguyen, Thang Viet Nguyen, Hieu Huy Pham, and Tuan Ngoc-Minh Nguyen. Deployment and validation of an ai system for detecting abnormal chest radiographs in clinical settings. *Frontiers in Digital Health*, page 130, 2022c.

Sam B Tran, Huyen TX Nguyen, Hieu H Pham, and Ha Q Nguyen. Transparency strategy-based data augmentation for bi-rads classification of mammograms. *arXiv preprint arXiv:2203.10609*, 2022.

Huyen TX Nguyen, Sam B Tran, Dung B Nguyen, Hieu H Pham, and Ha Q Nguyen. A novel multi-view deep learning approach for bi-rads and density assessment of mammograms. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2144–2148. IEEE, 2022d.

Ha Q Nguyen, Khanh Lam, Linh T Le, Hieu H Pham, Dat Q Tran, Dung B Nguyen, Dung D Le, Chi M Pham, Hang TT Tong, Diep H Dinh, et al. Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. *Scientific Data*, 9(1):429, 2022e.

Dat T Ngo, Hieu H Pham, Thao TB Nguyen, Hieu T Nguyen, Dung B Nguyen, and Ha Q Nguyen. Slice-level detection of intracranial hemorrhage on ct using deep descriptors of adjacent slices. *arXiv preprint arXiv:2208.03403*, 2022.

Hieu H Pham, Ha Q Nguyen, Hieu T Nguyen, Linh T Le, and Lam Khanh. An accurate and explainable deep learning system improves interobserver agreement in the interpretation of chest radiograph. *IEEE Access*, 10:104512–104531, 2022b.

Thao Nguyen, Tam M Vo, Thang V Nguyen, Hieu H Pham, and Ha Q Nguyen. Learning to diagnose common thorax diseases on chest radiographs from radiology reports in vietnamese. *Plos one*, 17(10):e0276545, 2022f.

Thao Nguyen, Hieu H Pham, Khiem H Le, Anh-Tu Nguyen, Tien Thanh, and Cuong Do. Detecting covid-19 from digitized ecg printouts using 1d convolutional neural networks. *PLoS One*, 17(11):e0277081, 2022g.

Quang-Huy Nguyen, Cuong Q Nguyen, Dung D Le, Hieu H Pham, and Minh N Do. Enhancing few-shot image classification with cosine transformer. *arXiv preprint arXiv:2211.06828*, 2022h.

Khiem H Le, Tuan V Tran, Hieu H Pham, Hieu T Nguyen, Tung T Le, and Ha Q Nguyen. Learning from multiple expert annotators for enhancing anomaly detection in medical image analysis. *IEEE Access*, 11:14105–14114, 2023.

Nguyen Tuan, Phi Nguyen, Dai Tran, Hung Pham, Quang Nguyen, Thanh Le, Hanh Van, Bach Do, Phuong Tran, Vinh Le, et al. Ensemble learning of myocardial displacements for myocardial infarction detection in echocardiography. *arXiv preprint arXiv:2303.06744*, 2023.

Thuy Dung Nguyen, Tuan Nguyen, Phi Le Nguyen, Hieu H Pham, Khoa Doan, and Kok-Seng Wong. Backdoor attacks and defenses in federated learning: Survey, challenges and future research directions. *arXiv e-prints*, pages arXiv–2303, 2023.

Van-Tuan Tran, Huy-Hieu Pham, and Kok-Seng Wong. Personalized privacy-preserving framework for cross-silo federated learning. *arXiv preprint arXiv:2302.12020*, 2023.

Nang Hung Nguyen, Duc Long Nguyen, Trong Bang Nguyen, Thanh-Hung Nguyen, Huy Hieu Pham, Truong Thao Nguyen, and Phi Le Nguyen. Cadis: Handling cluster-skewed non-iid data in federated learning with clustered aggregation and knowledge distilled regularization. *arXiv preprint arXiv:2302.10413*, 2023a.

Anh Duy Nguyen, Huy Hieu Pham, Huynh Thanh Trung, Quoc Viet Hung Nguyen, Thao Nguyen Truong, and Phi Le Nguyen. High accurate and explainable multi-pill detection framework with graph neural network-assisted multimodal data fusion. *arXiv preprint arXiv:2303.09782*, 2023b.

Tung-Trong Nguyen, Hieu H Pham, Phi Le Nguyen, Thanh Hung Nguyen, and Minh Do. Multi-stream fusion for class incremental learning in pill image classification. In *Proceedings of the Asian Conference on Computer Vision*, pages 4565–4580, 2022i.

Quan Nguyen, Hieu H Pham, Kok-Seng Wong, Phi Le Nguyen, Truong Thao Nguyen, and Minh N Do. Feddct: Federated learning of large convolutional neural networks on resource constrained devices using divide and co-training. *arXiv preprint arXiv:2211.10948*, 2022j.

Nang Hung Nguyen, Phi Le Nguyen, Thuy Dung Nguyen, Trung Thanh Nguyen, Duc Long Nguyen, Thanh Hung Nguyen, Huy Hieu Pham, and Thao Nguyen Truong. Feddrl: Deep reinforcement learning-based adaptive aggregation for non-iid data in federated learning. In *Proceedings of the 51st International Conference on Parallel Processing*, pages 1–11, 2022k.