Research Statement

ZHANG Zhiyuan
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828 5214; Email: zhiyuanzhang@smu.edu.sg
20 Dec 2024

Background

My research interests encompass the field of 3D computer vision, deep learning, and pattern recognition, with a focus on lightweight 3D point cloud analysis, rotation-invariant representation, multimodal learning for 3D point cloud, and 3D biometrics. With the proliferation of consumer-grade depth sensors, acquiring 3D data has become more intuitive and robust, resulting in numerous publicly available 3D datasets. This development has sparked growing interest in addressing scene understanding and analysis in the 3D domain. The aforementioned research areas are of great significance to both academia and industry as they form the basis of various applications, including 3D reconstruction, object detection, and autonomous driving, as well as robot navigation. My long-term research goal is to enable devices to comprehend everything in our 3D world.

Research Areas

<u>Lightweight 3D Point Cloud Analysis</u>
LiDAR is widely considered the most reliable sensor for autonomous robot and UAV navigation, thanks to its fast acquisition, high precision, and robustness to light variance. However, the raw data format, which is usually a large 3D point cloud, presents a challenge for real-time processing and deployment on edge devices. While many methods have been proposed in recent years, there is still a need for neural network compression, which can be achieved in two ways. The first approach is to compress at the algorithm level, using novel schemes such as lightweight and effective convolution operators, and new data representations to improve efficiency. This strategy has proven effective in my previous work, ShellNet [1], which proposes a lightweight convolution with only 0.5M trainable parameters and 1.5B FLOPs. Other methods usually have 3.5M-50M trainable parameters and take about 25B-30B FLOPs in the inference stage. Despite its compact size, ShellNet achieves the highest accuracy. The second solution is to compress at the architecture level, further reducing the size of the network through techniques such as knowledge distillation, network pruning, and binary networks. These topics are my future research interests with the goal to facilitate deployment onto ubiquitous MCU chips on edge devices.

<u>Rotation Invariant Representation for 3D Point Cloud Analysis</u>
While there has been recent progress in 3D point cloud deep learning, most prior works have focused on learning features that are invariant to translation and point permutation, with limited attention given to rotation invariance. However, rotation invariance is crucial for accurately recognizing and classifying objects in 3D point clouds, as objects can be viewed from different viewpoints, leading to variations in their orientation. Prior approaches have relied on rotation augmentation in training stages to alleviate the effects of rotation during testing. However, such a scheme is less effective given that 3D data has more degrees of

freedom. A rotation-invariant model is therefore critical to accurately recognize and classify objects in 3D point clouds. My previous work, RIConv [2], addressed this problem by transforming the original coordinates into rotation-invariant attributes, such as distance and angles, followed by binning and convolution. This approach resolved both rotation and point permutation invariance. However, it only considered local features, which could result in accuracy degradation. To overcome this limitation, GCAConv [3] was proposed by building a global context-aware convolution based on anchors and Local Reference Frames (LRF) to achieve rotation invariance. Although this approach showed improvement, LRF suffered from sign flipping problems in the x and y directions. To remove this uncertainty in LRF, I later proposed RIConv++ [4], which extracted informative features based on local reference axes and further improved accuracy to the same level as its translation-invariant counterpart. My goal for this research topic is not only to close the performance gap between rotation-invariant and non-rotation-invariant methods but also to surpass state-of-the-art methods while maintaining rotation invariance. To achieve this, a novel rotation invariant convolution operator called RISurConv [12] is proposed which captures the local surface structure and surpass both rotation-invariant and non-rotation-invariant methods by a large margin.

Multi-Modal Learning for Point Cloud Analysis

While deep learning has shown remarkable success in single-modal data analysis, combining multiple sources of information has the potential to further improve the accuracy and robustness of 3D point cloud analysis. Specifically, I aim to explore the synergistic effects of incorporating complementary data modalities such as color, texture, and different viewpoints, in addition to point cloud data. My recent work CVFNet [5] proposed multi-modal 3D object detection method used for autonomous driving by fusing different view features and point clouds. I will also propose approach involves developing new architectures that can effectively integrate multi-modal data and learn robust and discriminative features that are invariant to the variations in the input data. In addition, I plan to investigate the effectiveness of transfer learning between different modalities to reduce the need for large amounts of labeled data. The ultimate goal of my research on this topic is to advance the state-of-the-art in point cloud analysis by leveraging the complementary advantages of different data modalities through multi-modal learning, strengthening the applications in fields such as autonomous driving, robotics, and augmented reality.

3D Biometrics

Traditional biometric methods, such as fingerprint, facial, ear recognition [11], have been widely used, but they are vulnerable to spoofing and presentation attacks. 3D biometric analysis provides an additional layer of security by capturing the unique shape and surface characteristics of human body parts, such as face, hand, and ear, in a three-dimensional space. My research aims to develop novel 3D biometric algorithms that can accurately and efficiently extract and analyze the unique features of human body parts from 3D scans. I pioneered this direction by proposing a series of 3D dental identification methods [6-8]. To achieve efficient and effective identification, learning keypoint detection was proposed based on a novel informative 3D descriptor by designing signed angles [9] and histograms [10]. In future, I will investigate deep learning-based new methods for 3D feature representation, fusion, and feature matching, as well as data fusion to improve the development of accuracy and reliability of dental identification. I will also develop robust and accurate registration algorithms for handling the variations in pose and shape. My goal on this topic is to provide reliable and efficient solutions for forensic and medical applications with the aim for improved patient care and forensic investigations.

Selected Publications and Outputs

[1] Zhiyuan Zhang, Binh-Son Hua, Sai-Kit Yeung. “ShellNet: Efficient Point Cloud Convolutional Neural Networks using Concentric Shells Statistics”, International Conference on Computer Vision (ICCV Oral), 2019

[2] Zhiyuan Zhang, Binh-Son Hua, David W. Rosen, Sai-Kit Yeung. “Rotation Invariant Convolutions for 3D Point Clouds Deep Learning”, International Conference on 3D Vision (3DV), 204-213, 2019

[3] Zhiyuan Zhang, Binh-Son Hua, Wei Chen, Yibin Tian, Sai-Kit Yeung. “Global Context Aware Convolutions for 3D Point Cloud Understanding”, International Conference on 3D Vision (3DV Oral), 210-219, 2020

[4] Zhiyuan Zhang, Binh-Son Hua, Sai-Kit Yeung. “RIConv++: Effective Rotation Invariant Convolutions for 3D Point Clouds Deep Learning”, International Journal of Computer Vision (IJCV), 130(5): 1228-1243, 2022

[5] Jiaqi Gu, Zhiyu Xiang, Pan Zhao, Tingming Bai, Lingxuan Wang, Zhiyuan Zhang. “CVFNet: Real-time 3D Object Detection by Learning Cross View Features”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 568-574, 2022

[6] Zhiyuan Zhang, Sim Heng Ong, Xin Zhong, Kelvin W.C. Foong. “Efficient 3D Dental Identification via Signed Feature Histogram and Learning Keypoint Detection”, Pattern Recognition, 60, 189-204, 2016

[7] Xin Zhong, Zhiyuan Zhang. “3D Dental Biometrics: Automatic Pose-Invariant Dental Arch Extraction and Matching”, International Conference on Pattern Recognition (ICPR), 6524-6530, 2020

[8] Zhiyuan Zhang, Xin Zhong, Sim Heng Ong, Kelvin W.C. Foong. “An Efficient Partial Shape Matching Algorithm for 3D Tooth Recognition”, The 15th International Conference on BioMedical Engineering (ICBME), 785-788, 2013

[9] Zhiyuan Zhang, KangKang Yin, Kelvin W.C. Foong. “Symmetry Robust Descriptor for NonRigid Surface Matching”, Computer Graphics Forum (Proc. Pacific Graphics), 32(7), 355-362, 2013

[10] Zhiyuan Zhang, Sim Heng Ong, Kelvin W.C. Foong. “Improved Spin Images for 3D Surface Matching using Signed Angles”, IEEE International Conference on Image Processing (ICIP), 537-540, 2012

[11] Zhiyuan Zhang, Heng Liu. “Multi-view Ear Recognition Based on B-Spline Pose Manifold Construction”, The 7th World Congress on Intelligent Control and Automation (WCICA): 2416-2421, 2008

[12] Zhiyuan Zhang, Licheng Yang, Zhiyu Xiang. “RISurConv: Rotation Invariant Surface Attention-Augmented Convolutions for 3D Point Cloud Classification and Segmentation”, European Conference on Computer Vision (ECCV Oral), 93-109, 2024