Zhenyu Li

Personal Details:
Date of Birth: 5th May, 1999
Nationality: Chinese

Harbin City, Heilongjiang, 150001, China
(+86) 188-0041-9432
vgumypxt@gmail.com
Zhenyu Li | LinkedIn
Zhenyu Li | GitHub
Zhenyu Li | HomePage
Zhenyu Li | Google Scholar

Resourceful and future-forward post-graduate researcher with a keen desire to pursue research-focused PhD programme research fields on computer vision, especially for the 3D-related topics, to translate rich academic and industry experiences into state-of-the-art results; credited with master's degree in Computer Science & Technology from Harbin Institute of Technology, a C9 league (top-tier) institute of China.

Capable of leveraging strong programming skills to conduct in-depth research on object recognition and motion understanding in images, as well as building perceptual robotic and software systems based on the visual representations. Proven history of developing influential algorithms, models, software, and datasets to push the scientific frontiers of fundamental computer vision problems and visual learning systems. Eager to learn a self-contained account of computer vision and its underlying concepts, including the recent use of deep learning. Articulate communicator, possessing excellent presentation, writing, analytical, and critical thinking skills, with a meticulous attention to detail.

Selected Achievements

- Developed the first self-training framework for monocular 3D object detection on unsupervised domain adaptation, which solved the severe prediction shift caused by various imagery devices and significantly facilitated the application of Mono3D.
- Successfully realised the first multi-modal self-supervised system combining coupled camera and LiDAR data, which learnt spatially-aware visual representations for benefiting downstream 3D-related tasks.
- Resolved and further boosted AutoAlign performance by optimising large computational costs introduced by global attention for a learnable paradigm in fusing two modalities of Point clouds and RGB images for 3D object detection.
- Built and released a benchmark codebase on Github for Monodepth, receiving 400+ stars within 4 months.

Education & Credentials

Master of Science in Computer Science & Technology | Harbin Institute of Technology, China. Assessment 85.7, 2/260.

Bachelor of Science in Computer Science & Technology, 2021 | Harbin Institute of Technology, China. GPA 89.56, 66/261.

Research & Industry Experience

Didi Cargo, Beijing, China **2022 - Present**

Elite Research Intern

Conducting research, mainly basic research, on the autonomous driving system. Analysing the 'semi-supervised monocular 3D object detection', having a profound importance to academics and the industry.

- Substantially reducing the necessity for costly human annotations and enabling efficient use of the data acquired by driving automobiles, resulting in the improvement of perceptual capabilities of autonomous systems (ADAS) through the exact implementation of the research proposal.

SenseTime Research, Shanghai, China **2021 – 2022**

Research Intern, 2022

Analysed the problem for an industrial project, a collaboration between SenseTime autonomous driving group and the GAC group, a famous car manufacturer in China. Solved the need of applying a Mono3D model on a target dataset without labels. Mainly consider the images captured by different devices, in other words, the camera's intrinsic parameters. Authored a paper and proved the model's acceptable performance on the GAC required data. Continued research on the subject, independently afterwards, to reveal unexplained issues, such as training impediments, presence of object size bias, and non-consideration of camera FOV. Wrote a second paper on the topic with all the problems addressed. Honoured to experience the second paper recognised by the SenseTime group for use in the project with GAC.

- Explored unsupervised domain adaptation algorithms for monocular 3D object detection, resulting in a satisfactory model performance on the target domain; Received acceptance for the paper in ECCV 2022, as the first author.
- Achieved the deployment of the unsupervised domain adaptation algorithm for monocular 3D object detection in the industrial project with GAC Group.
- Studied multi-view monocular 3D object detection algorithms focusing on overlapping regions and drafted a paper as the second author to be accepted by ACM MM 2022.
- Explored domain generalization algorithms for monocular 3D object detection; completed a paper as the first author.

Perception Algorithm Development & Research Intern, 2021

Enhanced multi-object tracking (MOT) in the autonomous driving system (ADAS). Transformed the Sort, an algorithm, into DeepSort, an upgraded version of Sort algorithm. Added an appearance module by collecting a ReID dataset using photos and LiDAR points and training a ReID model. Obtained appearance-related representations to improve the association logic and develop C++ for optimising the coding in systems, allowing it to be used for industrial orientation. Received the approval to conduct further research activities in recognition of excellent study. Examined academic publications and performed practical research on multi-modal unsupervised pre-training to experiment and achieve effective results.

- Developed a ReID dataset based on the ground-truth system using both images and point clouds for the ADAS.
- Built and deployed the DeepSort multi-object tracking algorithm in the ADAS (C++), including importation of appearance representation from ReID model and adopting cascade association strategy; algorithm formed a patent for SenseTime.
- Successfully realised the first multi-modal self-supervised system combining coupled camera and LiDAR data, that learnt spatially-aware visual representations for benefit downstream 3D-related tasks.
- Conducted the research project under the supervision of Ang Li, Hongyang Li, Bolei Zhou, and Hang Zhao; submitted and received acceptance for the paper as the first author at AAAI 2022.
- Analysed multi-modal 3D object detection algorithms; submitted and gained acceptance for the paper as the second author at IJCAI 2022; the second paper as the second author accepted to ECCV 2022.

Publications

1. **Zhenyu Li**, Zehui Chen, Ang Li, Liangji Fang, Qinhong Jiang, Xianming Liu, Junjun Jiang, **"Unsupervised Domain Adaptation for Monocular 3D Object Detection via Self-Training"** ECCV 2022
   - Researched on the monocular camera is a cheap and practical setup for 3D object detection in ADAS.
   - Investigated the performance degradation (i.e., depth shift) of Mono3D detectors on the unseen dataset and proposed the first solution to Mono3D unsupervised domain adaptation.
   - Introduced the geometry-aligned multi-scale training strategy to disentangle the camera parameters and guarantee the geometry consistency of domains.
   - Designed a teacher-student paradigm to generate adaptive pseudo labels on the target domain.
   - Suggested the quality aware supervision, positive focusing training and dynamic threshold strategies to further facilitate model training.
   - Achieved remarkable performance on all evaluated datasets and surpassed supervised results on KITTI dataset.
2. Zehui Chen, **Zhenyu Li**, Shiquan Zhang, Liangji Fang, Qinhong Jiang, Feng Zhao, **"Deformable Feature Aggregation for Dynamic Multi-Modal 3D Object Detection"** ECCV 2022
   - Resolved and further boosted AutoAlign performance by optimising large computational costs introduced by global attention for a learnable paradigm in fusing two modalities of Point clouds and RGB images for 3D object detection.
   - Extended AutoAlign (IJCAI 2022 paper) to AutoAlignV2 with deformable multi-scale attention.
   - Highlighted the investigation of heterogeneous features in fusion-based 3D detection; adopted two effective augmentation strategies for fusion-based 3D object detection.
   - Achieved 72.4 NDS on nuScenes test leaderboard in the best model, attaining new, promising results among all published multi-modal 3D object detectors.
3. **Zhenyu Li**, Zehui Chen, Ang Li, Liangji Fang, Qinhong Jiang, Xianming Liu, Junjun Jiang, Bolei Zhou, Hang Zhao, **"SimIPU: Simple 2D Image and 3D Point Cloud Unsupervised Pre-Training for Spatial-Aware Visual Representations"** AAAI 2022
   - Proposed a simple, effective framework to adopt multi-modal self-supervised learning for better visual presentations, using tremendous data captured by vehicles daily, with an aim of effectively employing raw unlabelled data to benefit autonomous driving systems.
   - Developed a multi-modal contrastive learning framework, consisting of an intra-modal spatial perception module to learn a spatial-aware representation from point clouds and an inter-modal feature interaction module to transfer the capability of perceiving spatial information from the point cloud encoder to the image encoder.
   - Succeeded in obtaining desired results on three downstream tasks, including fusion-based 3D object detection, monocular 3D object detection, and monocular depth estimation.
4. Zehui Chen, **Zhenyu Li**, Shiquan Zhang, Liangji Fang, Qinhong Jiang, Feng Zhao, Bolei Zhou, Hang Zhao, **"AutoAlign: Pixel-Instance Feature Aggregation for Multi-Modal 3D Object Detection"** IJCAI 2022
   - Identified the challenges for object detection using either RGB images or the LiDAR point clouds, focusing on complementary and advantageous aspects of two data sources.
   - Proposed the AutoAlign dynamically fusing features in a fusion-based 3D detector; implemented constructive learning techniques to model a practical affinity map for cross-modal attention.
   - Analysed experimental results and realised 2.3 mAP and 7.0 mAP improvements on the KITTI and nuScenes datasets, respectively; Succeeded in reaching 70.9 NDS on the nuScenes testing leaderboard for the best model, accomplishing competitive performance among various sophisticated techniques used.
5. Zehui Chen, **Zhenyu Li**, Shiquan Zhang, Liangji Fang, Qinhong Jiang, Feng Zhao, **"Graph-DETR3D: Rethinking Overlapping Regions for Multi-View 3D Object Detection"** ACM MM 2022
   - Conducted intensive pilot experiments to quantify the objects located at different regions.

- Discovered the truncated instances, as the major constriction impeding the performance of DETR3D, a famous multi-view 3D object detector.
- Enhanced DETR3D into a spatial-graph version to facilitate model detection for truncated objects through sparse sampling in 3D space.
- Performed extensive experiments on the nuScenes dataset to demonstrate the effectiveness and efficiency of Graph-DETR3D; achieved 49.5 NDS on the nuScenes test leaderboard for the best model, attaining new, promising results as compared with various published image-view 3D object detectors.

6. **Zhenyu Li**, Zehui Chen, Ang Li, Liangji Fang, Qinhong Jiang, Xianming Liu, Junjun Jiang, **"Towards Model Generalization for Monocular 3D Object Detection"** Arxiv (May submit to CVPR2023)
    - Studied to resolve the unsolved issues in STMono3D, including considering the difficulty of data collection for the target domain, the performance degradation caused by the FOV difference, and the presence of object size bias.
    - Undertook the first study of the Mono3D domain generalisation and further analysed the disparity across cameras and the size bias among datasets.
    - Outperformed on all investigated datasets through DGMono3D technique and exceeded the SoTA unsupervised domain adaptation strategy by utilising statistical knowledge on target domain.

7. **Zhenyu Li**, Zehui Chen, Xianming Liu, Junjun Jiang, **"DepthFormer: Exploiting Long-Range Correlation and Local Information for Accurate Monocular Depth Estimation"** Arxiv
    - Implemented a new evaluation technique for monocular depth estimation and discovered an intriguing phenomenon: a model with a Transformer/convolution encoder outperforms on long-range/close-range depth estimation, respectively.
    - Suggested a parallel encoder architecture and a lateral fusion module that combines the most advantageous aspects of each, based on the aforementioned observation.
    - Experimented the KITTI, NYU, and SUN RGBD datasets to reveal the significantly exceeding the contemporary monocular depth estimation techniques using the proposed model, dubbed DepthFormer; obtained an exceptionally competitive score on the KITTI online depth estimation benchmark.

8. **Zhenyu Li**, Xuyang Wang, Xianming Liu, Junjun Jiang, **"BinsFormer: Revisiting Adaptive Bins for Monocular Depth Estimation"** Arxiv (Submitted to IEEE TIP)
    - Presented the first DETR-like decoder for monocular depth estimation in this study, as driven by MaskFormer.
    - Integrated a multiscale decoder structure to gain a full grasp of spatial geometry information and coarse-to-fine estimates of depth maps, while presenting an additional scene understanding query to increase estimation precision.
    - Conducted extensive testing on the KITTI, NYU, and SUN RGBD datasets revealing the BinsFormer outpacing contemporary monocular depth estimation techniques, with the benchmark ranking first on the extremely competitive KITTI online depth estimate website.

Projects

Monocular Depth Estimation Toolbox (https://github.com/zhyever/Monocular-Depth-Estimation-Toolbox) - Major contributor, Codebase, 400+ stars | Monocular-Depth-Estimation-Toolbox is an open-source monocular depth estimation toolbox based on PyTorch and mmSegmentation v0.16.0.

- Reproduced several famous supervised depth estimation methods, such as BTS, Adabins, and DPT.
- Utilised the toolbox to propose three different depth estimation methods (DepthFormer, BinsFormer, and LiteDepth), achieving engaging results on depth estimation tasks.

Awards & Honours

- **2nd Place**, Mobile AI & AIM 2022 Monocular Depth Estimation Challenge (ECCV2022 Workshop), Online, 2022
- **1st Place**, The KITTI Vision Benchmark Suite: Monocular Depth Estimation (BinsFormer), Online, 2022
- **1st Place**, The KITTI Vision Benchmark Suite: Monocular Depth Estimation (DepthFormer), Online, 2021
- **6th Place (Best Fidelity)**, Mobile AI 2021 Monocular Depth Estimation Challenge (CVPR2021 Workshop), Online, 2021

Presentations & Invited Talks

1. **IEEE/CAA JAS Symposium Series II, Intelligent Visual Perception in Smart City**, 2021
    - Introduced our work: "Enhancing Self-supervised Monocular Depth Estimation via Discrete Disparity and Uncertainty", "SimIPU: Simple 2D Image and 3D Point Cloud Unsupervised Pre-Training for Spatial-Aware Visual Representations", "DepthFormer: Exploiting Long-Range Correlation and Local Information for Accurate Monocular Depth Estimation", and "BinsFormer: Revisiting Adaptive Bins for Monocular Depth Estimation".

2. **IEEE Visual Communications and Image Processing (VCIP) 2022 Tutorial: 3D signal compression and processing**
    - Website: https://bychao100.github.io/blog/2022/vcip-tutorial/
    - Tutorial about depth estimation from monocular images and 3D object detection from cameras.