Qi Ma

Zürich, Switzerland 8057

+41 779968848 | qi.ma@vision.ee.ethz.ch | qimaqi.github.io | github.com/qimaqi | linkedin.com/in/qi-ma-27a655189/

Personal Profile

I am a PhD Candidate at Computer Vison Lab in ETH Zurich, Switzerland under the supervision of Prof. Dr. Ender Konukoglu and joint supervised by Prof. Dr. Luc Van Gool and Dr. Danda Pani Paudel in INSAIT, Soifa. My areas of interest include 3D vision, implicit representation, 3D segmentation, foundation model and LLM

Education

ETH Zurich, Switzerland

PhD in Computer Science — *April 2024 - Current*

- 3D Medical Segmentation: Large Vision Transformer on CT Segmentation
- Implicit function: Learning on Weight-Space, NeRF and its application
- 3DGS: Learning on Gaussian Space and its reasoning

ETH Zurich, Switzerland — *Grade: 5.95/6.0 ↑, top 10%*

MSc in Robotics and System Control — *Sept 2020 - Feb 2024*

- Master's thesis received **full grade 6.0** and Finalist for **ETH Outstanding Master's thesis**
- Teaching Assistant of course Probabilistic Artificial Intelligence
- Research Assistant of Computer Vision and Learning Group
- Researcher of student driverless team AMZ, **First design prize**

Karlsruhe Institut of Technology, Germany — *Grade: 1.43/1.0 ↓, top 5%*

Exchange student in Mechanical Engineering — *March 2019 - August 2019*

- Excellent exchange Scholarship

Shanghai Jiao Tong University, China — *Grade: 90.14/100 ↑, top 5%*

BSc of Science in Mechanical Engineering — *Sept 2016 - June 2020*

- Graduated with Distinction
- Bachelor Thesis **CPHS 2020 best student paper**

Projects

ShapeSplat & Gaussian-MAE — *ETH Zurich, INSAIT, Sofia*

A Large-scale Dataset of Gaussian Splats and Their Self-Supervised Pretraining — *May 2024 - Aug 2024*

- **Project Description:** 3D Gaussian Splatting (3DGS) has become the de facto method of 3D representation in many vision tasks. This calls for the 3D understanding directly in this representation space. To facilitate the research in this direction, we first build a large-scale dataset of 3DGS using the commonly used ShapeNet and ModelNet datasets. Our dataset ***ShapeSplat*** consists of 65K objects from 87 unique categories, whose labels are in accordance with the respective datasets. The creation of this dataset utilized the compute equivalent of **2 GPU years** on a TITAN XP GPU. We utilize our dataset for unsupervised pretraining and supervised finetuning for classification and segmentation tasks. To this end, we introduce ***Gaussian-MAE***, which highlights the unique benefits of representation learning from Gaussian parameters.
- **Supervisor:** Dr. Danda Pani Paudel, Prof. Ender Konukoglu. Prof. Luc Van Gool.

Implicit-Zoo — *ETH Zurich, INSAIT, Sofia*

A Large-Scale Dataset of Neural Implicit Functions for 2D Images and 3D Scenes — *Dec 2023 - May 2024*

- **Project Description:** We introduce "Implicit-Zoo": a large-scale dataset requiring thousands of GPU training days designed to facilitate research and development in this field. Our dataset includes diverse 2D and 3D scenes, such as CIFAR-10, ImageNet-1K, and Cityscapes for 2D image tasks, and the OmniObject3D dataset for 3D vision tasks. We ensure high quality through strict checks, refining or filtering out low-quality data. Using Implicit-Zoo, we showcase two immediate benefits as it enables to: (1) *learn token locations* for transformer models; (2) *directly regress* 3D cameras poses of 2D images with respect to NeRF models. This in turn leads to an *improved performance* in all three task of image classification, semantic segmentation, and 3D pose regression thereby unlocking new avenues for research.
- **Supervisor:** Dr. Danda Pani Paudel, Prof. Ender Konukoglu. Prof. Luc Van Gool.

Continuous Pose for Monocular Cameras — *ETH Zurich*

Master Thesis at Compuet Vision Lab(Subsequent work) — *Mar 2023 - Current*

- **Project Description:** Taking advantage of recent advancements in implicit function-based representations, the joint optimization of camera pose and scene structure becomes trivial. However, the potential of continuous motion characteristics of monocular cameras remains untapped. In this work we showcase the effectiveness of optimizing monocular camera poses as a **continuous function of time**. We exploit the proposed method in four diverse experimental settings, namely, (1) NeRF from noisy poses; (2) visual Simultaneous Localization and Mapping (vSLAM); (3) vSLAM with IMUs; and (4) NeRF from asynchronous Events. In all four settings, the proposed method performs significantly better than the compared baselines and the state-of-the-art methods.
- **Supervisor:** Dr. Ajad Chhatkuli, Dr. Danda Pani Paudel, Prof. Luc Van Gool.

Deformable Neural Radiance Fields using RGB and Event Cameras

ETH Zurich

Master Thesis at Compuet Vision Lab

Oct 2022 - Mar 2023

- **Project Description:** Modeling Neural Radiance Fields for fast-moving deformable objects from visual data alone is a challenging problem due to the high deformation and low acquisition rates. To address this problem, we propose to **use event cameras that offer very fast acquisition of visual change** in an asynchronous manner. Our proposed method uses the asynchronous stream of events and calibrated sparse RGB frames which jointly optimize these poses and the radiance field. This happens efficiently by leveraging the collection of events at once and actively sampling the events during learning. Experiments conducted on both realistically rendered graphics and real-world datasets demonstrate a significant benefit of the proposed method over the compared baseline.
- **Accomplishment: ICCV 2023** accepted.
- **Supervisor:** Dr. Ajad Chhatkuli, Dr. Danda Pani Paudel, Prof. Luc Van Gool.

Human Pose Estimation from Egocentric Social Interaction Videos

ETH Zurich

Student Project of course: Virtual Humans

Feb 2022 - June 2022

- **Responsibilities:** Analyzing the egocentric pose estimation model, Modifying the https://github.com/qimaqi/TCMR$_{R}ELEASETCMRmodelandco$
- **Supervisor:** Dr. Siwei Zhang.
- **Accomplishment:** The work achieved amazing performance compared to the https://github.com/facebookresearch/you2meYOU2ME original work, which formulates camera wearer pose estimation as a classification task. Lastly, By fitting the https://smpl.is.tue.mpg.de/SMPL model based on estimated keypoints and gained smooth and accurate results compared to running shape estimation directly. The performance of our method ranked **3rd in the Egobody competition**.Results https://github.com/qimaqi/VH$_{P}roj_{p}ublichere.$

Active Depth Sensing with an Event Camera

ETH Zurich

Semester Project at Robotics and Perception Group

Feb 2021 - Aug 2021

- **Responsibilities:** Designing the calibration method for event-only SL systems, processing the data and designing the experiments.
- **Project Description:** I propose a fast, accurate, and event-only calibration method for event-based structured light (SL) systems. Unlike other current calibration which relies on both events and images, our method utilizes only events (no frames are required during the entire process), I can unlock the full potential of high-resolution event cameras for high-speed, low-power SL system.
- **Supervisor:** Dr. Manasi Muglikar, Prof. Davide Scaramuzza.
- **Accomplishment:** Our first event-only calibration method is not only 16 times faster than state-of-the-art but also improves calibration accuracy by 43% due to accurately modeling the projector's distortion. Results https://github.com/qimaqi/Calibrating-an-Event-based-Structured-Light-Systemhere.

On the Robustness of Local Feature Inversion Techniques

ETH Zurich

Student Project of course: 3D Vision

March 2021 - June 2021

- **Responsibilities:** Designing the Inverting Network and Colorization, conducting experiments.
- **Project Description:** I built an inverting network and colorization network which can reconstruct highly-detailed images with realistic color information from learned local features such as https://arxiv.org/abs/1712.07629SuperPoint and https://arxiv.org/abs/1906.06195R2D2. Besides, I conducted exhaustive experiments towards the performance of three reconstruction network backbones, InvNet, UNet and UNet++, with different input feature sparsity levels on the learned features.
- **Supervisor:** Dr. Mihai Dusmanu.
- **Accomplishment:** I showed that not only it is feasible to reconstruct the original image from learned features, but the reconstruction result is far accurate in terms of both image integrity and texture details. Result https://github.com/qimaqi/3dv$_{p}ytorchhere.$

Multimodal Driving Data Fusion, Multi-task learning, 3D Object Detection.

ETH Zurich

Student Project of course: Deep learning for autonomous driving

March 2021 - June 2021

- **Responsibilities:** Developing the algorithm solving the tasks, testing it and also further improving it.
- **Project Description:** In the first task I compensate the egocentric motion of the car using information from IMU to get undistorted lidar points. In the second task I implemented the Multi-Task Learning (MTL) architectures for dense semantic segmentation and monocular depth estimation. In the last task I built a 2-stage 3D object detector to detect vehicles in autonomous driving scenes.
- **Accomplishment:** Full grade in the task1, 16.75/20 ↑in task2 and 10.25/12 ↑in task3. Result https://github.com/qimaqi/DLAD$_{e}xercises_{p}ublichere.$

Work Experience

Computer vision Intern

Zurich

Huawei Zurich Research Center

Mar 2023 - Sep 2023

- Developing algorithms for artificial-based video stabilization.
- Data collection, evaluation and further improve the algorithms.
- Accelerating the algorithms and Deploying it on mobile device
- **Technical Skills:** Pytorch, Git, Docker.

Teaching Assistant

ETH, Zurich

Probabilistic Artificial Intelligence

Sept 2022 - Feb 2023

- Designing a code assignment that allowed students to implement a Bayesian neural network with various variational inference methods on the corrupted MNIST dataset, including dropout, ensemble, MCMC and Bayes backprop.
- Developing https://github.com/qimaqi/BNN$_{S}egdemorunningtheBayesianneuralnetworkforsegmentationtaskonCamVidDataset$
- **Technical Skills:** Git, Docker, IPython, Gpytorch, Pytorch.

Research Assistant *ETH, Zurich*

Project: Human Pose Estimation from Egocentric Social Interaction Videos *Feb 2022 - June 2022*

- Data collection using multiple Kinect and HoloLens, data calibration and generating keypoints using https://github.com/CMU-Perceptual-Computing-Lab/openposeOpenpose.
- Incorrect Human keypoints cleaning using 3D box generated by https://github.com/zju3dv/EasyMocapEasyMocap and manual checking.
- Modifying the TCMR model for better Human pose estimation with egocentric pose information.
- **Technical Skills:** Kinects, HoloLens,Openpose, EasyMocap.

Software programmer *ETH, Zurich*

Formula Studnet Driverless Team: Academic Motorsports Club Zurich

https://driverless.amzracing.ch/en/home(AMZ) *Sept 2021 - Aug 2022*

- Implementing a sequential sensor-fusion based pipeline https://www.youtube.com/watch?v=6RJvNTTBE6g(video) designed for effective utilization of multi-modal data from the camera, lidar and output accurate 3D cones classes and location.
- Developing the KPI tool using vehicle RTK and self-developed GPS marking device for evaluating 3D cone localization accuracy. Modifying and training deep learning network YOLOv5 and implementing it in real-time using C++ and TensorRT for formula student competitions. Deploying model on Android application using JAVA. More details of the work https://github.com/qimaqi/AMZ$_{W}orkhere.Exploringthepossibilityofusingeventcamerasensorsetup.DevelopingEvent-YOLO$
- **Technical Skills:** Interface of Hesai Lidar,Basler Camera, and RTK GPS, Ubuntu, ROS, YOLOv5, TensorRT, JAVA.