Publications

I am a year-3 PhD student at CUHK CSE under the supervision of Prof. Pheng Ann Heng. My previous research aims at endowing the machines to sense and understand the visual world in the following areas: **(1) Multi-modal Learning**, especially sign language recognition and trajectory prediction. **(2) 3D Vision**, especially 3D understanding and 6D pose estimation. My representative publications are as follows:

[1] Traj-MAE: Masked Autoencoders for Trajectory Prediction
Hao Chen*, **Jiaze Wang***, Kun Shao, Furui Liu, Chenyong Guan, Jianye Hao, Guangyong Chen, Pheng-Ann Heng. ***ICCV 2023***

[2] PointPatchMix: Point Cloud Mixing with Patch Scoring
Yi Wang*, **Jiaze Wang***, Jinpeng Li, Zixu Zhao, Guangyong Chen, Anfeng Liu and Pheng-Ann Heng. ***AAAI 2024***

[3] TripletMix: Triplet Data Augmentation for 3D Understanding
Jiaze Wang*, Yi Wang*, Ziyu Guo, Renrui Zhang, Donghao Zhou, Guangyong Chen, Anfeng Liu, Pheng-Ann Heng. ***In submission***

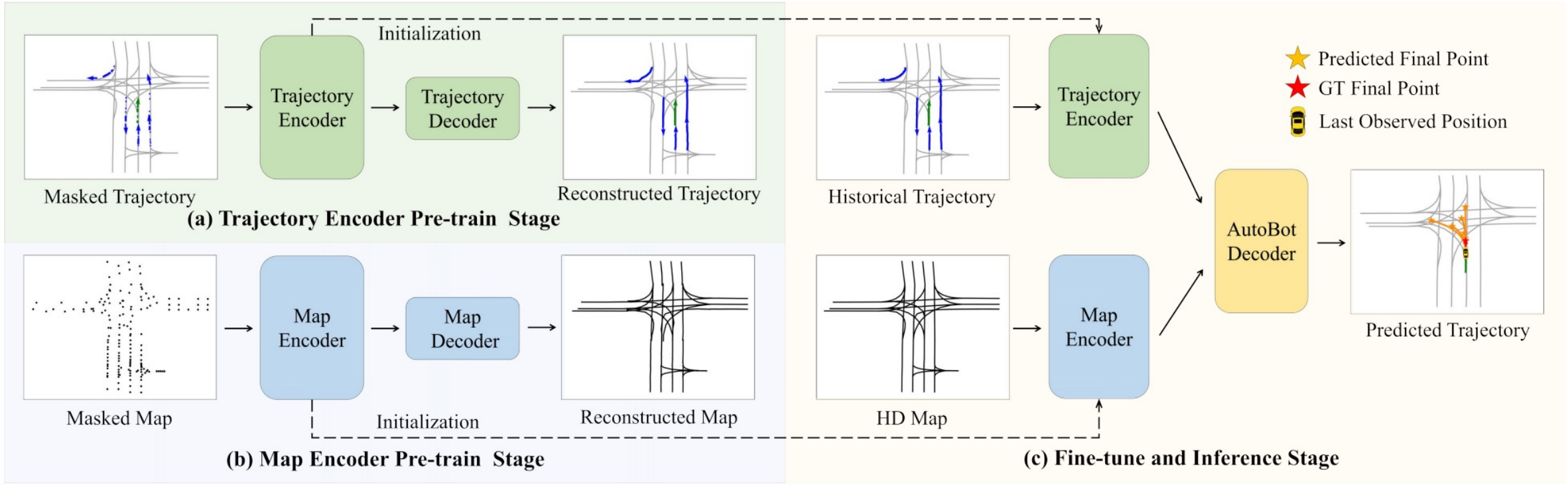
[4] SignVTCL: Multi-Modal Continuous Sign Language Recognition Enhanced by Visual-Textual Contrastive Learning
Hao Chen*, **Jiaze Wang***, Ziyu Guo, Jinpeng Li, Donghao Zhou, Bian Wu, Chenyong Guan, Guangyong Chen, and Pheng-Ann Heng. ***In submission***

[5] SFANet: Spatial-Frequency Attention Network for Weather Forecasting
Jiaze Wang, Hao Chen, Hongcan Xu, Jinpeng Li, Bowen Wang, Kun Shao, Furui Liu, Huaxi Chen, Guangyong Chen, and Pheng-Ann Heng. ***In submission***

[6] Object-centric Multiple Object Tracking
Zixu Zhao, **Jiaze Wang,** Max Horn, Yizhuo Ding, Tong He, Zechen Bai, Bing Shuai, Dominik Zietlow, Carl-Johann Simon-Gabriel, Zhuowen Tu, Thomas Brox, Bernt Schiele, Yanwei Fu, Tianjun Xiao, Francesco Locatello, Zheng Zhang. ***ICCV 2023***

[7] Category-Level 6D Object Pose Estimation via Cascaded Relation and Recurrent Reconstruction Network
Jiaze Wang*, Kai Chen*, and Qi Dou. ***IROS 2021***

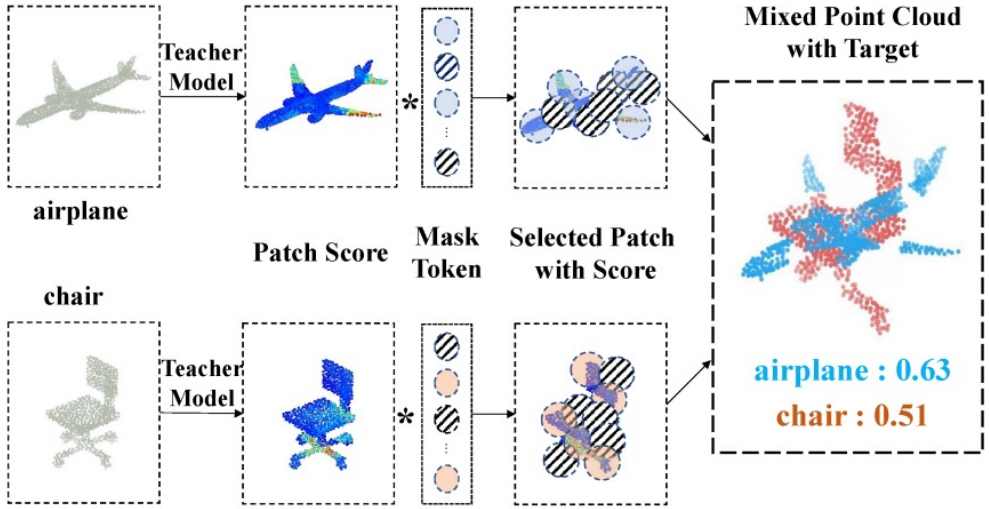
Trajectory Prediction



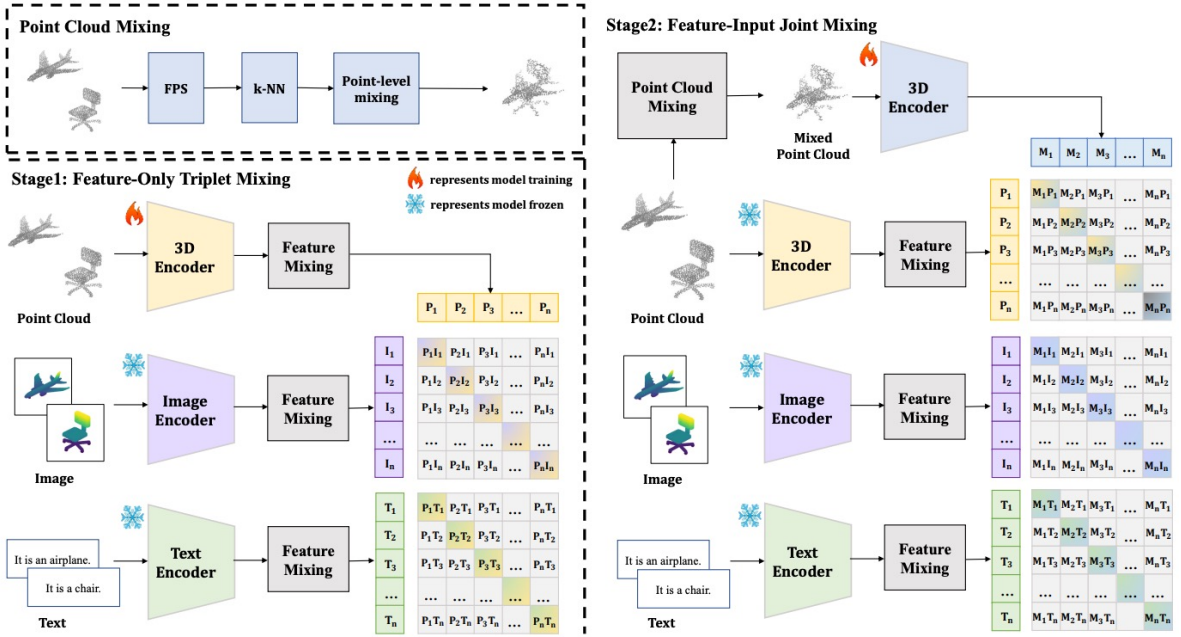
[1] Overview of **Traj-MAE**. Traj-MAE is mainly composed of three stages: (a) Trajectory encoder pre-train stage with continual trajectory masking and reconstruction strategies. (b) Map encoder pre-train stage with continual map masking and reconstruction strategies. (c) Fine-tune and inference stage where the encoders are initialized by the pre-trained models' parameters.

[1] Traj-MAE: Masked Autoencoders for Trajectory Prediction
Hao Chen*, **Jiaze Wang***, Kun Shao, Furui Liu, Chenyong Guan, Jianye Hao, Guangyong Chen, Pheng-Ann Heng.
ICCV 2023

3D Data Augmentation



[1]. Illustration of generating mixed data using **PointPatchMix**. Given two point clouds, PointPatchMix processes them at the patch level, with each patch comprising 32 points. A pre-trained teacher model scores each patch based on self-attention mechanism. Then, the mixed point cloud consists of patches selected by mask tokens and the new ground truth is generated.

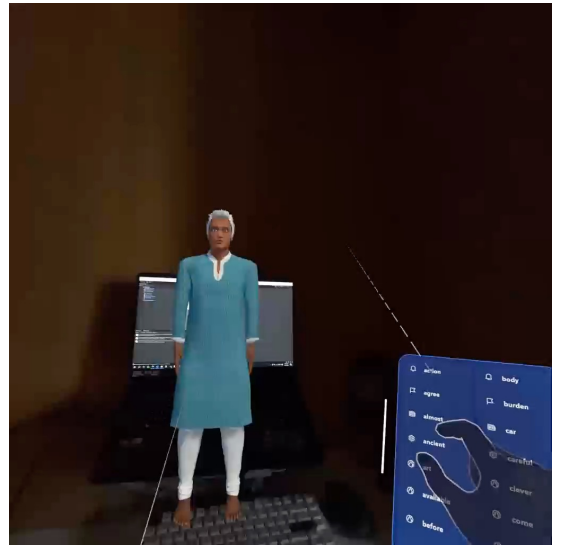


[2] The overall scheme of **TripletMix**. TripletMix consists of two stages. In the first stage, the 3D encoder is trainable, while the image and text encoders are pre-trained and frozen. Feature embeddings are extracted for contrastive learning with the 3D features. In the second stage, the trained 3D encoder initializes a new trainable 3D encoder. The other encoders remain frozen. Two input point clouds are mixed using FPS, k-NN, and point-level mixing, then fed into the new 3D encoder. The mixed features are used for contrastive learning with the features from the other encoders to align representations of all three modalities.

[2] PointPatchMix: Point Cloud Mixing with Patch Scoring
Yi Wang*, **Jiaze Wang***, Jinpeng Li, Zixu Zhao, Guangyong Chen, Anfeng Liu and Pheng-Ann Heng.
AAAI 2024

[3] TripletMix: Triplet Data Augmentation for 3D Understanding
Jiaze Wang*, Yi Wang*, Ziyu Guo, Renrui Zhang, Donghao Zhou, Guangyong Chen, Anfeng Liu, Pheng-Ann Heng
In submission

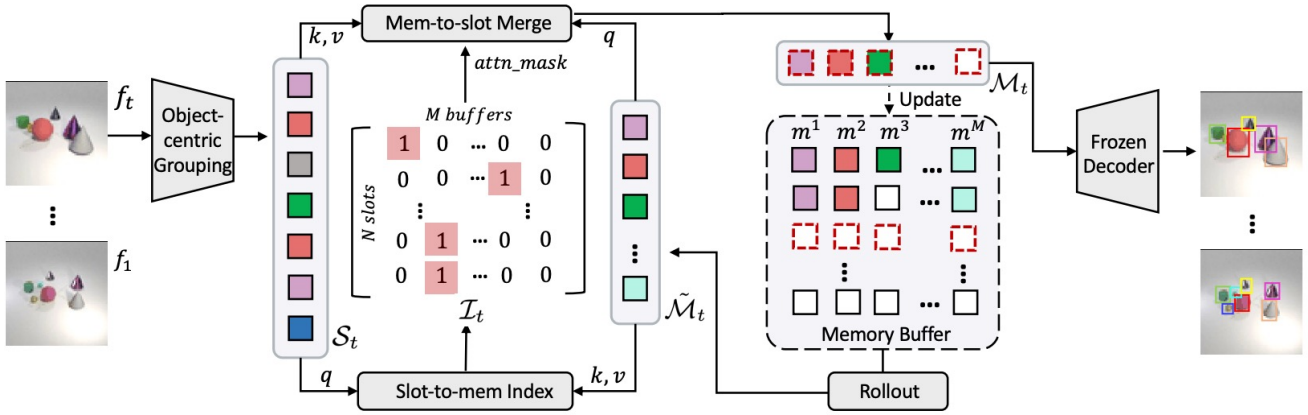
Sign Language Recognition (Ongoing)

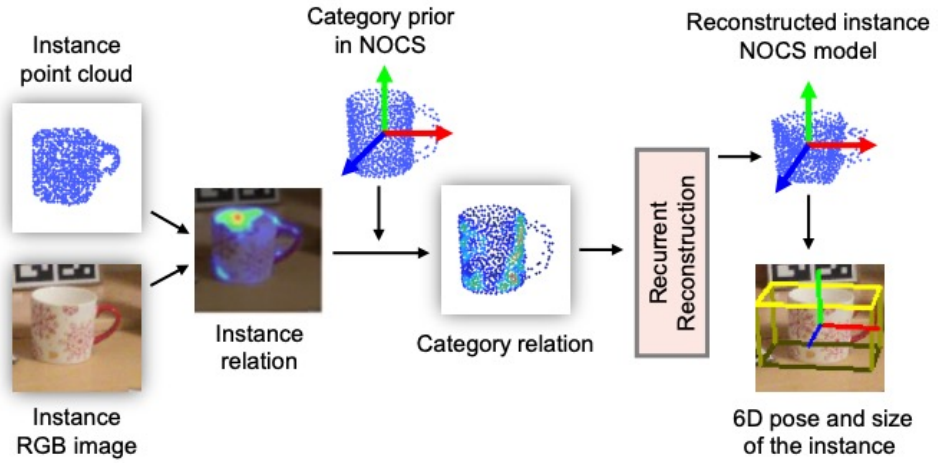


[4] SignVTCL: Multi-Modal Continuous Sign Language Recognition Enhanced by Visual-Textual Contrastive Learning

Hao Chen*, **Jiaze Wang***, Ziyu Guo, Jinpeng Li, Donghao Zhou, Bian Wu, Chenyong Guan, Guangyong Chen, and Pheng-Ann Heng. ***In submission***

Object Tracking and Pose estimation





[6] **OC-MOT** . It consists of two main modules. i) An index-merge module that adapts object-centric slots into detection results via two steps. First, index each slot into memory buffers by a learnable index matrix It indicating all the slot-tomemory assignments. Second, merge slots assigned to the same buffer by recalculating the attention weights masked by It backwards. ii) A object memory module that improves temporal consistency by rolling historical state forwards for object association. For MOT evaluation, we decode the detection results to masks or bounding boxes via a frozen decoder in the object-centric grouping module.

[7] Our proposed **category-level 6D pose estimation via cascaded relation and recurrent reconstruction networks**. The networks are mainly composed of two networks: (1) A cascaded relation network to exploit the relation of between RGB images and point clouds, and the relation between instance features and category features. (2) An recurrent reconstruction network for canonical shape reconstruction from coarse to fine.

[6] Object-centric Multiple Object Tracking

Zixu Zhao, **Jiaze Wang,** Max Horn, Yizhuo Ding, Tong He, Zechen Bai, Bing Shuai, Dominik Zietlow, Carl-Johann Simon-Gabriel, Zhuowen Tu, Thomas Brox, Bernt Schiele, Yanwei Fu, Tianjun Xiao, Francesco Locatello, Zheng Zhang. ***ICCV 2023***

[7] Category-Level 6D Object Pose Estimation via Cascaded Relation and Recurrent Reconstruction Network

Jiaze Wang*, Kai Chen*, and Qi Dou. ***IROS 2021***