Enze Xie

Phone: (+86) 18721073081 **Email:** xieenze@hku.hk

Homepage: https://xieenze.github.io

Highlights

- My research is computer vision and deep learning, *e.g.* object detection and segmentation in 2D and 3D.
- I am supervised by Prof. Ping Luo at HKU MMLab. Currently I am having an internship in NVIDIA, working with Zhiding Yu, Jose M. Alvarez, Sanja Fidler and Anima Anandkumar.
- My work PolarMask was selected as **CVPR 2020 Top-10 Influential Papers**.
- I co-developed OpenSelfSup (1.3k+ github star), which is a popular self-supervised learning toolbox.

Education

The University of Hong Kong, Hong Kong 2019 – Now
PhD: Computer Vision *Supervisor: Prof. Ping Luo.*

Tongji University, Shanghai 2016 – 2019
Master: Computer Science

Nanjing University of Aeronautics and Astronautics, Nanjing 2012 – 2016
Bachelor: Aircraft Manufacturing

Publications

Publication: **10** first/co-first papers (*e.g.* TPAMI, CVPR, ICCV, ECCV). Citation: **800+**
Google Scholar: https://scholar.google.com/citations?user=42MVVPgAAAAJ&hl=zh-CN
Github: https://github.com/xieenze

Journal

[1] PolarMask++: Enhanced Polar Representation for Single-Shot Instance Segmentation and Beyond
Enze Xie, Wenhai Wang, Mingyu Ding, Ruimao Zhang, Ping Luo
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021.

[2] PAN++: Towards Efficient and Accurate End-to-End Spotting of Arbitrarily-Shaped Text
Wenhai Wang, **Enze Xie**, Xiang Li, Ding Liang, Ding Liang, Zhibo Yang, Tong Lu, Chunhua Shen
IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021.

Conference

[1] SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers
Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, Ping Luo
Neural Information Processing Systems (NeurIPS), 2021

[2] DetCo: Unsupervised Contrastive Learning for Object Detection
Enze Xie*, Jian Ding*, Wenhai Wang, Xiaohang Zhan, Hang Xu, Zhenguo Li, Ping Luo
IEEE International Conference on Computer Vision (ICCV), 2021.

[3] Segmenting Transparent Objects in the Wild with Transformer
Enze Xie, Wenjia Wang, Wenhai Wang, Peize Sun, Hang Xu, Ding Liang, Ping Luo
International Joint Conference on Artificial Intelligence (IJCAI), 2021.

[4] Segmenting Transparent Objects in the Wild
Enze Xie, Wenjia Wang, Wenhai Wang, Mingyu Ding, Chunhua Shen, Ping Luo
European Conference on Computer Vision (ECCV), 2020.

[5] PolarMask: Single Shot Instance Segmentation with Polar Representation
Enze Xie*, Peize Sun*, Xiaoge Song*, Wenhai Wang, Ding Liang, Chunhua Shen, Ping Luo
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[6] Scene Text Detection with Supervised Pyramid Context Network
Enze Xie*, Yuhang Zang*, Shuai Shao, Gang Yu, Cong Yao, Guangyao Li
Association for the Advancement of Artificial Intelligence (AAAI), 2019

[7] Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions
Wenhai Wang, **Enze Xie**, Xiang Li, Deng-Ping Fan, Kaitao Song, ..., Ping Luo, Ling Shao
IEEE International Conference on Computer Vision (ICCV), 2021.

[8] Watch Only Once: An End-to-End Video Action Detection Framework
Shoufa Chen, Peize Sun, **Enze Xie**, Chongjian Ge, Jiannan Wu, Lan Ma, Jiajun Shen, Ping Luo
IEEE International Conference on Computer Vision (ICCV), 2021.

[9] What Makes for End-to-End Object Detection?
Peize Sun, Yi Jiang, **Enze Xie**, Wenqi Shao, Zehuan Yuan, Changhu Wang, Ping Luo
International Conference on Machine Learning (ICML), 2021.

[10] Scene Text Image Super-Resolution in the Wild
Wenjia Wang*, **Enze Xie***, Xuebo Liu, Wenhai Wang, Ding Liang, Chunhua Shen, Xiang Bai
European Conference on Computer Vision (ECCV), 2020.

[11] Differentiable Hierarchical Graph Grouping for Multi-Person Pose Estimation
Sheng Jin, Wentao Liu, **Enze Xie**, Wenhai Wang, Chen Qian, Wanli Ouyang, Ping Luo
European Conference on Computer Vision (ECCV), 2020.

[12] AE TextSpotter: Learning Visual and Linguistic Representation for Ambiguous Text Spotting
Wenhai Wang, Xuebo Liu, Xiaozhong Ji, **Enze Xie**, Ding Liang, ..., Chunhua Shen, Ping Luo
European Conference on Computer Vision (ECCV), 2020.

[13] Efficient and Accurate Arbitrary-Shaped Text Detection with Pixel Aggregation Network
Wenhai Wang*, **Enze Xie***,Xiaoge Song, Yuhang Zang, Tong Lu, Gang Yu, Chunhua Shen
IEEE International Conference on Computer Vision (ICCV), 2019.

[14] Shape Robust Text Detection with Progressive Scale Expansion Network
Wenhai Wang*, **Enze Xie***, Xiang Li, Wenbo Hou, Tong Lu, Gang Yu, Shuai Shao
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019

Under Submission

[1] Panoptic SegFormer
Zhiqi Li, Wenhai Wang, **Enze Xie**, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, Tong Lu, Ping Luo
arXiv:2109.03814

[2] CycleMLP: An MLP-like Architecture for Dense Prediction
Shoufa Chen, **Enze Xie**, Chongjian Ge, Ding Liang, Ping Luo
arXiv:2107.10224

[3] Unsupervised Pretraining for Object Detection by Patch Reidentification
Jian Ding*, **Enze Xie***, Hang Xu, Chenhan Jiang, Zhenguo Li, Ping Luo, Gui-Song Xia
arXiv:2103.04814, submitted to IEEE Trans. PAMI

[4] Improving Monocular Visual Odometry Using Robustly Learned Depth
Wei Yin, Libo Sun, **Enze Xie**, Zhengrong Li, and Changming Sun
Submitted to IEEE Trans. Robotics

[5] TransTrack: Multiple-Object Tracking with Transformer
Peize Sun, Yi Jiang, Rufeng Zhang, **Enze Xie**, Jinkun Cao, Xinting Hu, and Tao Kong et al.
arXiv:2012.15460

Experience

NVIDIA 2021.03 – Now

Research Intern (with Zhiding Yu, Jose M. Alvarez, Sanja Fidler and Anima Anandkumar)

- **Vision Transformer.** I develop a semantic segmentation algorithm based on Transformer, which is simple, efficient yet very powerful. It also shows excellent zero shot robustness. Submitted 1 paper to NeurIPS21.
- **Autonomous driving.** I am doing 3D object detection -> tracking -> forecasting based on six cameras.

Facebook, Apply Machine Learning Team 2020.05 – 2020.07

Project Collaborator (with Guan Pang)

- **Optical Character Recognition.** I develop a pipeline that can use limited data with less human annotation, while achieving strong performance. It can save many costs for the company. Submit 1 paper to ICCV21.

Huawei Noah's Ark Lab , AI Theory Group 2020.06 – 2021.2

Research Intern (with Hang Xu and Zhenguo Li)

- **Self-supervised learning.** I develop an algorithm that can pre-train on massive unlabeled data, and benefits many downstream tasks such as object detection. This algorithm is used in HUAWEI autonomous driving team. 1 paper accepted by ICCV21 and 1 paper submitted to TPAMI.
- **Transformers.** I develop a semantic segmentation algorithm based on Transformer. Accepted by IJCAI21.

SenseTime, General Model Group 2019.07 – 2020.03

Research Intern (with Ding Liang)

- Work on instance segmentation and human pose estimation. 1 paper accepted by CVPR2020 (Oral) and 4 paper accepted by ECCV2020.

Megvii, Detection Group 2018.04 – 2019.07

Research Intern (with Gang Yu and Shuai Shao)

- Apply detection algorithms to products *e.g.* car/license/pedestrian detection.
- Work with Dr. Gang Yu on scene text detection. 3 papers accepted by AAAI19, CVPR19 and ICCV19.

eBay Traffic Team 2017.07 – 2017.12

Big Data Develop Intern

- Use big data tools(*e.g.* Hadoop/Spark/Flume/Kafka/Kudu) to process thousands of TB of data in eBay.

Awards and Honors

- National Artificial Intelligence Challenge (NAIC) 2020 , Remote Sensing Semantic Segmentation Task, **1st Place, 1,000,000 RMB** 2020
- OpenImage 2019 Instance Segmentation, **1st Place** 2019
- ICDAR2019 Arbitrary-Shaped Text Detection, **1st Place** 2019
- ICDAR 2019 Large-scale Street View Text Detection, **2nd Place** 2019
- Outstanding Master Thesis Award, Tongji University 2019
- HKU Postgraduate Scholarship 2019-now
- HK & China Gas Co Ltd Postgraduate Prize 2021

Popular Projects

OpenSelfSup: Self-Supervised Learning Toolbox and Benchmark

Github (1.2k star): `https://github.com/open-mmlab/OpenSelfSup`

- OpenSelfSup is an open source unsupervised representation learning toolbox based on PyTorch.

Shape Robust Text Detection with Progressive Scale Expansion Network CVPR'19

Github (1k star): `https://github.com/whai362/PSENet`

- PSENet is a practical algorithm to detect scene text with arbitrary shape. It is used in many companies, e.g. HUAWEI, ByteDance, SenseTime and MEGVII.

PolarMask: Single Shot Instance Segmentation with Polar Representation CVPR'20 & TPAMI'21

Github (730 star): `https://github.com/xieenze/PolarMask`

- A novel method to represent instance with polar coordinate. CVPR20's Top-10 Influential Papers.

Academic Service

- CVPR2019, ICCV2019 Student Volunteer
- Conference Reviewer for NeurIPS, CVPR, ICCV, ECCV, AAAI, IJCAI, ACCV, WACV
- Journal Reviewer for T-MM, NeuroComputing, TNNLS, PR