Siwei Yang

swyang.ac@gmail.com | laos-y.github.io | github.com/LAOS-Y/

EDUCATION

University of California, Santa Cruz — Santa Cruz, CA
Ph.D. in Computer Science — *Jun. 2023 – Present*
- Advisor: Prof. Cihang Xie

Tongji University — Shanghai
B.Eng. in Computer Science — *Sep. 2017 – Jul. 2021*

EXPERIENCE

Research Summer Intern — Baltimore, MD
Johns Hopkins University (JHU), CCVL — *Jul. 2022 – Dec. 2022*
- Under Supervision of Prof. Alan Yuille and Prof. Hang Zhao.
- Worked on **Weakly-supervised instance segmentation**.
- One paper in submission to **CVPR 2023**.

Visiting Research Assistant — Hong Kong
HKUST, MMLab — *Jul. 2021 – Jul. 2022*
- Under Supervision of Prof. Dan Xu.
- Worked on **Multi-task Dense Prediction**.
- One paper accepted by **AAAI 2023**.

Research Intern — Beijing
MEGVII Research, Base Detection — *Jun. 2020 – Jun. 2021*
- Participated in the development and maintenance of Cvpods.
- Implemented MEGVII-proposed ThunderNet and CrowdDetRCNN in Cvpods.
- Added various features in Cvpods.
- Deep understanding and rich experience in **Object Detection** and **Semantic Segmentation**.

Undergraduate Research Assistant — Shanghai
Tongji University — *Feb. 2019 – Jun. 2021*
- Under the supervision of Prof. Yin Wang.
- Worked on semantic segmentation.
- One paper published at **CVPR 2020** the 1st Agriculture-Vision Challenge.
- **Best undergraduate** award in **NeurIPS 2020** SpaceNet 7 challenge (Team leader).

PUBLICATIONS

AQA-Bench: An Interactive Benchmark for Evaluating LLMs' Sequential Reasoning Ability in Algorithmic Environments
Siwei Yang*, Bingchen Zhao*, Cihang Xie.
In submission to *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track* (**NeurIPS 2024**).
TL;DR: This paper presents AQA-Bench, a benchmark for evaluating LLMs' sequential reasoning abilities via interactive environments requiring model executing algorithms such as Binary search, DFS, and BFS. Our find includes **(1)** the inverse scaling between model sizes and performance, **(2)** the nuanced impact of naive in-context examples due to over-fitting in ICL, **(3)** weak models failing mainly due to incapability of starting well and **(4)** impressive improvement from a few given predecessor steps following the optimal policy.

WebAgent-90K: A Large-Scale Dataset for Fine-Tuning Agent for Automatic Web Browsing Tasks
Bingchen Zhao*, **Siwei Yang***, Cihang Xie.
In submission to *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track* (**NeurIPS 2024**).

TL;DR: In this paper we present WebAgent-90K, web-interaction dataset with around 90K tasks collected via Evol-Instruct and an automated web agent based on GPT-4V, which can be used for training an automated web agent with open-sourced VLM. The Llava-v1.5 we finetuned with WebAgent-90K yielded similar performance as GPT-4V on Webvoyager.

HQ-Edit: A High-Quality Dataset for Instruction-based Image Editing
Mude Hui*, **Siwei Yang***, Bingchen Zhao, Yichun Shi, Heng Wang, Peng Wang, Yuyin Zhou, Cihang Xie
In submission to *European Conference on Computer Vision 2024* (**ECCV 2024**).
TL;DR: This study introduces HQ-Edit, a high-quality instruction-based image editing dataset with around 200K edits. Unlike prior approaches relying on human feedback, we devise a scalable data collection pipeline leveraging self-instruct with advanced foundation models, namely GPT-4V and DALL-E 3.

AsyInst: Asymmetric Affinity with DepthGrad and Color for Box-Supervised Instance Segmentation
Siwei Yang, Longlong Jing, Junfei Xiao, Hang Zhao, Alan Yuille, Yingwei Li.
Preprint.
TL;DR: Due to the optimization problem of the former symmetric pairwise affinity loss, it is only compatible with color affinity but not with other modalities. Our method alleviates this issue by introducing asymmetry, which not only makes it compatible with depth gradient affinity but also improves the performance with color affinity.

Contrastive Multi-Task Dense Prediction
Siwei Yang, Hanrong Ye, Dan Xu.
The Thirty-Seventh AAAI Conference on Artificial Intelligence (**AAAI 2023**).
TL;DR: We discover that in a multi-task model, task-specific features follow a cross-task contrastive distribution, e.g. pixels with the same semantic label have similar features for depth estimation. Therefore, we devise a regularization method that can improve multi-task performance by enhancing this distribution.

XCon: Learning with Experts for Fine-grained Category Discovery
Yixin Fei, Zhongkai Zhao, **Siwei Yang**, Bingchen Zhao.
British Machine Vision Conference 2022 (**BMVC 2022**) Oral.
TL;DR: Learning to do category discovery within a fine-grained dataset is challenging, we present a method that learns to do so by partitioning the dataset into k sub-groups, and shows improved performance on several fine-grained datasets.

Reducing the feature divergence of RGB and near-infrared images using Switchable Normalization
Siwei Yang*, Shaozuo Yu*, Bingchen Zhao*, Yin Wang. (* means equal contribution)
IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 2020 (**CVPRW 2020**).
TL;DR: Instance normalization reduces feature divergence between modalities while batch normalization keeps the discriminative distribution. Thus, segmentation models achieve better performance by utilizing both kinds of normalization.

PROJECTS

Cvpods   github.com/Megvii-BaseDetection/cvpods
A versatile and efficient codebase for multiple computer vision tasks: classification, segmentation, detection, keypoints, self-supervised learning, etc. The aim of Cvpods is to achieve **efficient experiments management** and **smooth task-switching**.
- Added various features: OHEM, Soft-NMS, Cluster-NMS, Dataset registering, Visualization, etc.
- Added various models: ShuffleNet, SNet, ThunderNet, CrowdDetRCNN, etc.
- Built CI from scratch including linting, unit test and model training.

AgriVision   github.com/LAOS-Y/AgriVision
Code for AGRICULTURE-VISION 2020 (CVPRW 2020). Various backbones(ResNet, IBN-Net, ResNeXt) and decoders(UNet, DlinkNet, FPN, DeeplabV3+) are supported. Final ranking at the **8th place**. One workshop paper published at **CVPRW2020**.

- Implemented features including AMP, multi-GPU training, solver, data augmentation, etc.
- Added support for various models and loss functions.
- Designed and conducted part of the validation experiments.
- Summarized our work into the published paper.

2-Stage EGFR   github.com/LAOS-Y/2Stage-EGFR

A 2-stage lung cancer classifier based on CT images using Pytorch. A one-stage nodule detector with FPN is used to predict bboxes of nodules. Then Bbox-cropped CT patches of nodules are fed to a classifier for further classification.

Awards and Achievement

No. 2 of **MICCAI 2023** BraTS Challenge, Brain Metastases Segmentation Track.	2023
Co-organizer of the 1st OOD-CV workshop at **ECCV 2022**.	2022
Best Undergraduate award in **NeurIPS 2020** SpaceNet7 Challenge.	2020
No.8 out of 39 teams in **CVPR 2020** Agriculture-Vision Challenge Track.	2020
No.27 out of 134 teams in CCF Multi-race Face Recognition Challenge.	2019
Second prize at National Olympiad in Informatics in Provinces.	2015

Academic Services

Reviewer for CVPR 2022, ECCV 2022, WACV 2023, ICRA 2023, NeurIPS (2021, 2022).