Sheng Zhang

971+052-7536-696 | [Github](#) | [Website](#) | [sheng.zhang@mbzuai.ac.ae](#) | [eyisheng@outlook.com](#)

EDUCATION

Mohamed Bin Zayed University of Artificial Intelligence June 2023 – Present
Research Assistant, UAE
Supervisor: [Prof. Kun Zhang](#)

Mohamed Bin Zayed University of Artificial Intelligence August 2021 – May 2023
Master of Science (Graduate Research Assistant), UAE
Supervisor: [Prof. Salman Khan](#); Co-supervisor: [Dr. Zhiqiang Shen](#)

Tongji University September 2016 – June 2021
Bachelor of Software Engineering, Shanghai, China

RESEARCH INTERESTS

○ Open-World Generalization: visual category discovery (with vision-language models), open-world learning.
○ Latent Variable Models: VAE, generative model, identifiability of latent variables with causal discovery.

PUBLICATIONS

1. Towards Realistic Zero-Shot Classification via Self Structural Semantic Alignment.
   Sheng Zhang, *Muzammal Naseer, Guangyi Chen, Zhiqiang Shen, Salman Khan, Kun Zhang, Fahad Khan.* AAAI 2024 (**Oral**). [Paper Code](#).

2. PromptCAL: Contrastive Affinity Learning via Auxiliary Prompts for Generalized Novel Category Discovery.
   Sheng Zhang, *Salman Khan, Zhiqiang Shen, Muzammal Naseer, Guangyi Chen, Fahad Khan.* CVPR 2023. [Paper Code](#).

3. A Channel Attention Based Deep Neural Network for Automatic Metallic Corrosion Detection.
   Sheng Zhang, *Xinling Deng, Yumin Lu, Shaozheng Hong, Zhengyi Kong, Yongli Peng, Ye Luo.* Journal of Building Engineering, 2021. [Paper](#).

RESEARCH EXPERIENCE

Realistic Zero-Shot Classification with Vision-Language Learning Jan. 2023 – June. 2023
My individual research supervised by Prof. Salman Khan, Dr. Zhiqiang Shen, Prof. Fahad Khan. *MBZUAI, UAE*

➤ `Intro`: Formulated a novel problem, *Realistic Zero-Shot Classification*, which relaxes the *ideal vocabulary* assumption of the ground-truth target label set. Proposed a semantic structural alignment method in our framework with calibrated image-category alignments by leveraging Large Language Models.

• Adapted and reproduced multiple previous strong baselines to our setting for performance comparisons.

• Developed and benchmarked our method as SOTA performance on six generic and fine-grained datasets evaluated with transductive/inductive, unsupervised/semi-supervised, and out-of-vocabulary setups. In the standard setup, our method surpasses CLIP by over 20% absolute accuracies.

Generalized Novel Category Discovery June 2022 – Dec. 2022
My individual research and master thesis supervised by Prof. Salman Khan *MBZUAI, UAE*

➤ `Intro`: Proposed a visual prompt-based contrastive affinity learning framework to address the generalized category discovery problem, which aims to categorize both known and novel classes with known class annotations. Our method can learn semantic discriminative clusters via contrastive learning on diffused affinity graphs.

• Designed and implemented a novel visual prompt regularization technique to enhance backbone semantic discriminativeness. Demonstrated its superiority over naive visual prompt tuning.

• Achieved SOTA performance on seven challenging benchmarks including fine-grained StanfordCars and CUB, significantly surpassing previous methods, e.g., with nearly 11% cluster accuracy on CUB and 9% on ImageNet-100. Conducted further evaluations in transductive/inductive and few-annotation scenarios.

Weakly-Supervised Semantic Segmentation on Medical Images Jan. 2021 – June 2021
My bachelor thesis supervised by Dr. Ye Luo. *Tongji University, China*

➤ `Intro`: Developed an adversarial self-supervised approach to address semi-supervised domain adaptation problem.

- Designed a multi-scale patch rotation prediction pretext to enhance the representation transferability on downstream low-annotation segmentation tasks, surpassing RotNet and JigSaw pretexts by over 3% average IoU.
- Evaluated methods by comparing with RotNet and Jigsaw representation learning baselines in the semi-supervised setting of various labeling ratios and in the domain-shit scenario.

Solution to Mask-Scale Variations in Medical Image Segmentation   July 2020 – Jan. 2021
My collaborative research for the MICCAI2020 TN-SCUI challenge supervised by Dr. Ye Luo. Tongji University, China
- ➢ `Intro`: Designed a differentiable two-stage network to address the problem of large variations in target scales.
- Conduct experimental ablations on UNet architecture with various techniques architectural improvements, including channel and gate attention mechanism, and D-LinkNet attention on Pancreas-CT scans dataset.

Metallic Corrosion Detection System   Mar. 2020 – Jan. 2021
My collaborated research supervised by Dr. Ye Luo.   *Tongji University, China*
- ➢ `Intro`: Implemented an automatic framework to detect multi-level metallic corrosion based on channel attention.
- Adapted squeeze-and-excitation network for spatial metallic corrosion detection with channel-wise explainability.
- Conducted method comparisons, and ablations on spatial and channel attention mechanisms and explained metallic corrosion spatially with CAM visualization technique on channel-wise aggregated data.

PROJECT EXPERIENCE

Class-agnostic Open-Set Object Detection   Sep. 2021 – June. 2022
Advised by Prof. Salman Khan.   *MBZUAI, UAE*
- ➢ `Intro`: Detecting all objects including unannotated categories with a class-agnostic detector by addressing the inherent weakness of closed-set detectors, *i.e.*, the issue of false negative penalties.
- Enhanced the openness of detectors through a copy-paste data augmentation strategy by over +10% on AR@1000.
- Designed and implemented an uncertainty-based box-jittering method to refine detector localization prediction.

Partially-Supervised Medical Image Segmentation   Mar. 2021 – June 2021
Course project advised by Dr. Mohammad Yaqub.   *MBZUAI, UAE*
- ➢ `Intro`: Segmenting semantic organs on 3D MRI-scan datasets with only partial voxel-wise annotations. For each scan, only one type of organs is annotated. The learning goal is to segment all types of organs.
- Designed and implemented an adapted CutMix with Spatial Priori with +2.4% average IoU before post-processing, and partially-supervised contrastive loss with $\sim$ +0.3% average IoU compared with state-of-the-art DoDNet.

A Study on Visual Relationship Detection with Large-Scale Pre-trained Models   Mar. 2022 – May. 2022
Course project of NLP   *MBZUAI, UAE*
- ➢ `Intro`: Studying properties of visual relationship datasets with large-scale pre-trained models, *e.g.*, CLIP, BERT.
- Finetuned pretrained CLIP to the visual relationship detection task.
- Implemented an effective semi-supervised framework for visual relationship plausibility detection.

Semi-Supervised Speech Sentiment Analysis with Pre-trained Models   Mar. 2022 – May. 2022
Course project of NLP advised by Dr. Shady Shehata.   *MBZUAI, UAE*
- ➢ `Intro`: Exploring pre-trained language/audio models for semi-supervised speech sentiment analysis.
- Reproduced from scratch and adapted a semi-supervised speech sentiment analysis paper without released codes.
- Adapted the framework with Wav2Vec 2.0 model and benchmarked on SAVEE and IEMOCAP datasets.

ACADEMIC SERVICES

○ Delivered a presentation on our work *Towards Realistic Zero-Shot Classification* at AAAI'24
○ Serving as a reviewer at NIPS, CVPR, ICCV, ECCV, IJCV, TIP, AISTATS, ACM Computing Surveys
○ Serving as a conference volunteer at AAAI'24, EMNLP'23

SKILLS

Language: Chinese (native), English (fluent): TOEFL iBT 106
Programming: Python, JAVA, Oracle, Git
Libraries: Pytorch, Scikit-learn, Matplotlib, Pandas, NumPy, Scipy, NetworkX, ...
Other DL Frameworks: MMDetection, MMCV, Pytorch-Lightning, timm
Others: C++, Matlab, Scrapy, PL/SQL, MySQL, Hadoop (Hive, HBase, Impala, Spark/PySpark)