SEONGHEON PARK

Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea

Phone : (82-10) 8221-9574 ⬦ E-mail : sam121796@yonsei.ac.kr ⬦ https://seongheon-park.netlify.app/

RESEARCH INTERESTS

Computer Vision, Machine Learning, Image and Video Understanding, Vision-Language Models, Out-of-Distribution Detection, Long-Tailed Learning, Learning with Limited Labeled Data, Open-World Learning

EDUCATION

Master of Electrical and Electronic Engineering, Yonsei University *Sep. 2021 - Aug. 2023*
(Advisor : Prof. Kwanghoon Sohn)
Cumulative GPA : 4.0/4.0

Bachelor of Electrical and Electronic Engineering, Yonsei University *Mar. 2015 - Aug. 2021*
Cumulative GPA : 3.60/4.0, Major GPA : 3.90/4.0, Upper GPA : 3.91/4.0

WORKING PAPERS

1. **Seongheon Park**, Hyuk Kwon, and Kibok Lee (2024), "Towards Imbalanced Open-World Semi-Supervised Learning", *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*

PUBLICATIONS

1. Hanjae Kim, Jiyoung Lee, **Seongheon Park**, and Kwanghoon Sohn (2023), "Hierarchical Visual Primitive Experts for Compositional Zero-Shot Learning", *IEEE/CVF International Conference on Computer Vision (ICCV)*

2. Minsu Kim, Seungryong Kim, Jungin Park, **Seongheon Park**, and Kwanghoon Sohn (2023), "PartMix: Regularization Strategy to Learn Part Discovery for Visible-Infrared Person Re-identification", *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*

3. **Seongheon Park**, Hanjae Kim, Minsu Kim, and Kwanghoon Sohn (2023), "Normality Guided Multiple Instance Learning for Weakly Supervised Video Anomaly Detection", *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*

4. Dahye Kim, Jiyoung Lee, Jungin Park, **Seongheon Park**, and Kwanghoon Sohn (2023), "Language-free Training for Zero-shot Video Grounding", *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*

CONFERENCE PRESENTATIONS

1. **Seongheon Park**, Hanjae Kim, Minsu Kim, and Kwanghoon Sohn, "Normality Guided Multiple Instance Learning for Weakly Supervised Video Anomaly Detection", *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, Hawaii, Jan. 2023, Oral Presentation*

PROJECTS

1. **Project with Korea Institute of Science and Technology (KIST)** *Jan. 2022 - Aug. 2023*

   - Project Title : Deep Identification and Tracking of Missing Persons in Heterogeneous CCTV
   - Covers : Object Detection, Object Tracking, Video Action Classification, Person Re-Identification, Anomaly Detection

2. **Project with Ministry of Science and ICT, Mid-Level Research** *Jan. 2022 - Aug. 2023*

   - Project Title : Development of Complex Situational Awareness and Prediction Technology through Multi-Modal Data Fusion and Social Artificial Intelligence
   - Covers : Multi-Modal Learning, Video Understanding

PATENTS

- 1 US Patent, 1 Domestic Patent, 1 Software Registration

HONORS AND SCHOLARSHIPS

Honors

- Honors Student, Yonsei University   *2018*
- Honors Student, Yonsei University   *2016*

Scholarships

- Teaching Assistant Scholarship, Yonsei University   *2022*
- Research Assistant Scholarship, Yonsei University   *2021 - 2023*
- Brain Korea 21 (BK21) Scholarship , National Research Foundation of Korea   *2021 - 2023*
- Future Vehicula Technology Scholarship,
  Inter-University Alliance for Future Vehicula Technology   *2021 - 2023*
- Academic Excellence Scholarship, Yonsei University   *2019*

WORK AND RESEARCH EXPERIENCES

Graduate Researcher, Machine Learning Lab.   *Sep. 2023 - Present*
(Advisor : Prof. Kibok Lee)

- Open-World Learning, Out-of-Distribution Detection
- (Self-Supervised) Representation Learning
- Long-Tailed Learning
- Learning with Limited Labeled Data

Research Assistant, Digital Image Media Lab.   *Sep. 2021 - Aug. 2023*
(Advisor : Prof. Kwanghoon Sohn)

- Open-World Learning, Out-of-Distribution Detection
- Image/Video Anomaly Detection
- Vision-Language Models
- Cross-Domain Person Re-Identification
- Learning with Limited Labeled Data

Undergraduate Research Assistant, Digital Image Media Lab.   *Jan. 2021 - Aug. 2021*
(Advisor : Prof. Kwanghoon Sohn)

- Image and Video Understanding
- Human-Object Interaction Detection

Undergraduate Research Assistant, Multi-Dimensional Insight Lab.   *Jan. 2019 - Mar. 2019*
(Advisor : Prof. Sanghoon Lee)

- Deep learning-based hand pose estimation

Personnel Administration, R.O.K. Logistics Command (Military Service).   *Jul. 2019 - Jan. 2021*

- Managed soldiers and carried out administrative work

TEACHING EXPERIENCES

- Introduction to Digital Labs (Undergraduate)   *Fall 2022*
- Digital Signal Processing (Undergraduate)   *Spring 2022*
- Image Coding (Graduate)   *Spring 2022*

SELECTED COURSEWORK

Electrical and Electronic Engineering :   Data structure, linear algebra, probability and random variables, random process, special topics for deep learning, probabilistic robotics, reinforcement learning, optimization theory, information theory, pattern recognition, artificial intelligence

Others :   Deep learning for computer vision, mathematical statistics, Bayesian statistics, advanced deep learning

SKILL SET

Language and Tools :   Python, Matlab, C++, R, LaTeX

Libraries :   Pytorch, TensorFlow, Keras, Scikit-Learn, Numpy, Jupyter, etc.

PUBLICATION ABSTRACT

1. Hanjae Kim, Jiyoung Lee, **Seongheon Park**, and Kwanghoon Sohn (2023), "Hierarchical Visual Primitive Experts for Compositional Zero-Shot Learning",   *IEEE/CVF International Conference on Computer Vision (ICCV)*

- Compositional zero-shot learning (CZSL) aims to recognize unseen compositions with prior knowledge of known primitives (attribute and object). Previous works for CZSL often suffer from grasping the contextuality between attribute and object, as well as the discriminability of visual features, and the long-tailed distribution of real-world compositional data. We propose a simple and scalable framework called Composition Transformer (CoT) to address these issues. CoT employs object and attribute experts in distinctive manners to generate representative embeddings, using the visual network hierarchically. The object expert extracts representative object embeddings from the final layer in a bottom-up manner, while the attribute expert makes attribute embeddings in a top-down manner with a proposed object-guided attention module that models contextuality explicitly. To remedy biased prediction caused by imbalanced data distribution, we develop a simple minority attribute augmentation (MAA) that synthesizes virtual samples by mixing two images and oversampling minority attribute classes. Our method achieves SoTA performance on several benchmarks, including MIT-States, C-GQA, and VAW-CZSL. We also demonstrate the effectiveness of CoT in improving visual discrimination and addressing the model bias from the imbalanced data distribution.

2. Minsu Kim, Seungryong Kim, Jungin Park, **Seongheon Park**, and Kwanghoon Sohn (2023), "PartMix: Regularization Strategy to Learn Part Discovery for Visible-Infrared Person Re-identification",   *IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*

- Modern data augmentation using a mixture-based technique can regularize the models from overfitting to the training data in various computer vision applications, but a proper data augmentation technique tailored for the part-based Visible-Infrared person Re-IDentification (VI-ReID) models remains unexplored. In this paper, we present a novel data augmentation technique, dubbed PartMix, that synthesizes the augmented samples by mixing the part descriptors across the modalities to improve the performance of part-based VI-ReID models. Especially, we synthesize the positive and negative samples within the same and across different identities and regularize the backbone model through contrastive learning. In addition, we also present an entropy-based mining strategy to weaken the adverse impact of unreliable positive and negative samples. When incorporated into existing part-based VI-ReID model, PartMix consistently boosts the performance. We conduct experiments to demonstrate the effectiveness of our PartMix over the existing VI-ReID methods and provide ablation studies.

3. **Seongheon Park**, Hanjae Kim, Minsu Kim, and Kwanghoon Sohn (2023), "Normality Guided Multiple Instance Learning for Weakly Supervised Video Anomaly Detection", *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*

- Weakly Supervised Video Anomaly Detection (wVAD) aims to distinguish anomalies from normal events based on video-level supervision. Most existing works utilize Multiple Instance Learning (MIL) with ranking loss to

tackle this task. These methods, however, rely on noisy predictions from a MIL-based classifier for target instance selection in ranking loss, degrading model performance. To overcome this problem, we propose Normality Guided Multiple Instance Learning (NG-MIL) framework, which encodes diverse normal patterns from noise-free normal videos into prototypes for constructing a similarity-based classifier. By ensembling predictions of two classifiers, our method could refine the anomaly scores, reducing training instability from weak labels. Moreover, we introduce normality clustering and normality guided triplet loss constraining inner bag instances to boost the effect of NG-MIL and increase the discriminability of classifiers. Extensive experiments on three public datasets (ShanghaiTech, UCF-Crime, XD-Violence) demonstrate that our method is comparable to or better than existing weakly supervised methods, achieving state-of-the-art results.

4. Dahye Kim , Jiyoung Lee, Jungin Park, **Seongheon Park**, and Kwanghoon Sohn (2023), "Language-free Training for Zero-shot Video Grounding", *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*

- Given an untrimmed video and a language query depicting a specific temporal moment in the video, video grounding aims to localize the time interval by understanding the text and video simultaneously. One of the most challenging issues is an extremely time- and cost-consuming annotation collection, including video captions in a natural language form and their corresponding temporal regions. In this paper, we present a simple yet novel training framework for video grounding in the zero-shot setting, which learns a network with only video data without any annotation. Inspired by the recent language-free paradigm, i.e. training without language data, we train the network without compelling the generation of fake (pseudo) text queries into a natural language form. Specifically, we propose a method for learning a video grounding model by selecting a temporal interval as a hypothetical correct answer and considering the visual feature selected by our method in the interval as a language feature, with the help of the well-aligned visual-language space of CLIP. Extensive experiments demonstrate the prominence of our language-free training framework, outperforming the existing zero-shot video grounding method and even several weakly-supervised approaches with large margins on two standard datasets.