Youngtaek Oh

Ph.D. Student, Korea Advanced Institute of Science and Technology (KAIST)

✉ youngtaek.oh@kaist.ac.kr ⌂ https://ytaek-oh.github.io in LinkedIn ⌾ Github 🎓 Google Scholar

Research Interest

My research centers on developing data-efficient and robust recognition systems that address limitations and biases in image and video data. I am currently focusing on enhancing vision-language models to improve compositional reasoning and multi-modal understanding for more robust and comprehensive recognition systems.

Keywords: Data-Efficient Learning, Multi-Modal Learning, Vision-Language Models

Education

Korea Advanced Institute of Science and Technology (KAIST)	Daejeon, South Korea
Ph.D. in Electrical Engineering	Sep. 2021 – Present
○ Co-Advisors: Prof. In So Kweon and Prof. Junmo Kim	
M.S. in Electrical Engineering	Mar. 2019 – Feb. 2021
○ Advisor: Prof. In So Kweon	
Korea University	Seoul, South Korea
B.S. in Electrical Engineering	Mar. 2015 – Feb. 2019
○ GPA: 4.40/4.50	

Publications (*equally contributed authors; †corresponding authors)

Workshop and Preprints

[W2] Exploring the Spectrum of Visio-Linguistic Compositionality and Recognition
Youngtaek Oh, Pyunghwan Ahn, Jinhyung Kim, Gwangmo Song, Soonyoung Lee†, In So Kweon†, Junmo Kim†
in **CVPRW 2024**: 'What is Next in Multimodal Foundation Models?' (**MMFM**) Workshop

[W1] Technical Report: Retrieval-based Data Discovery and Fusion for Zero-shot Image Captioning
Youngtaek Oh, Jae Won Cho, Dong-Jin Kim, In So Kweon†, Junmo Kim†
in **CVPRW 2023**: 'New Frontiers for Zero-shot Image Captioning Evaluation' (**NICE**) Workshop
2nd place in Zero-Shot Image Captioning Challenge in NICE Workshop

Peer-Reviewed Conferences and Journals

[C6] Preserving Multi-Modal Capabilities of Pre-trained VLMs for Improving Vision-Linguistic Compositionality
Youngtaek Oh, Jae Won Cho, Dong-Jin Kim, In So Kweon†, Junmo Kim†
in **EMNLP 2024**: Conference on Empirical Methods in Natural Language Processing
Oral presentation

[J1] Empirical study on using Adapters for debiased Visual Question Answering
Jae Won Cho, Dawit Mureja Argaw, **Youngtaek Oh**, Dong-Jin Kim, In So Kweon
in **CVIU 2023**: Computer Vision and Image Understanding (IF=4.3)

[C5] Self-Sufficient Framework for Continuous Sign Language Recognition
Youngjoon Jang, **Youngtaek Oh**, Jae Won Cho, Myungchul Kim, Dong-Jin Kim, In So Kweon, Joon Son Chung
in **ICASSP 2023**: IEEE International Conference on Acoustics, Speech and Signal Processing
Oral presentation, Top 3% Paper Recognition

[C4] Signing Outside the Studio: Benchmarking Background Robustness for Continuous Sign Language Recognition
Youngjoon Jang, **Youngtaek Oh**, Jae Won Cho, Dong-Jin Kim, Joon Son Chung, In So Kweon
in **BMVC 2022**: British Machine Vision Conference

[C3] DASO: Distribution-Aware Semantics-Oriented Pseudo-Label for Imbalanced Semi-Supervised Learning
Youngtaek Oh, Dong-Jin Kim, In So Kweon
in **CVPR 2022**: IEEE/CVF Conference on Computer Vision and Pattern Recognition
Finalist, Qualcomm Inovation Fellowship Korea, 2022

[C2] KSL-Guide: A Large-scale Korean Sign Language Dataset Including Interrogative Sentences for Guiding the Deaf and Hard-of-Hearing
Soomin Ham, Kibaek Park, Youngjoon Jang, **Youngtaek Oh**, Seokmin Yun, Sukwon Yoon, Chang Jo Kim, Han-Mu Park, In So Kweon
in **FG 2021**: IEEE International Conference on Automatic Face and Gesture Recognition

[C1] SideGuide: A Large-scale Sidewalk Dataset for Guiding Impaired People
Kibaek Park*, **Youngtaek Oh***, Soomin Ham*, Kyungdon Joo*, Hyokyoung Kim, Hyoyoung Kum, In So Kweon
in **IROS 2020**: International Conference on Intelligent Robots and Systems

Work Experience

Research Intern	Sep. 2023 – Feb. 2024
LG AI Research	Seoul, South Korea

- Worked on vision-language compositionality; Outcome: Paper [W2] and Software [S3]

Graduated Researcher	Mar. 2021 – Aug. 2021
Korea Advanced Institute of Science and Technology (KAIST)	Daejeon, South Korea

Projects

Developing and Demonstrating Innovative Products Based on Public Demand Nov. 2021 – Present
funded by the Ministry of Science and ICT (MSIT), Korea

- Consortium: KAIST, Hanbat National University, Miru Systems, Hanulsoft, Datamaker, Daejeon Transportation
- Objective: Develop a real-time system for masking and unmasking personal information in public CCTV services.
- Role: Designed a deep steganography algorithm to conceal and reveal personal information (face and license plate).

Korean Sign Language Dataset for AI Interpretation May 2020 – Dec. 2020
funded by the National Information Society Agency (NIA), Korea

- Consortium: KAIST, Korea Association of the Deaf, Testworks, Korea Nazarene University, EQ4ALL
- Objective: Establish a large-scale Korean Sign Language dataset for real-world use in sign language recognition.
- Role: Developed a continuous sign language recognition model compatible with Korean Sign Language datasets.
- Outcome: **Publications** ([C2], [C4], [C5]), **dataset** (KSL-Guide ↗, *available to Koreans only*)

Korean Sidewalk Image Dataset for AI Assistance to the Visually Impaired May 2019 – Dec. 2019
funded by the National Information Society Agency (NIA), Korea

- Consortium: KAIST, Korea Spinal Cord Injury Association, Testworks, SelectStar, DTWORESOURCE
- Objective: Construct a dataset for sidewalk environments to enhance mobility rights for individuals with disabilities.
- Role: Designed the sidewalk obstacle dataset and validated it using object recognition models.
- Outcome: **Publications** ([C1]), **dataset** (SideGuide ↗, *available upon request: application ↗*)

Honors and Awards

Top 3% Recognition Certificates, ICASSP 2023	June 2023
Second place in the 2023 NICE Challenge at the NICE Workshop, CVPR 2023 ($5,000)	May 2023
Finalist, Qualcomm Inovation Fellowship Korea	Nov. 2022
Outstanding Reviewer Award, ECCV 2022	Oct. 2022
National Scholarship for Science and Engineering *(full tuition, merit-based)*	Mar. 2017 – Feb. 2019

Softwares

Open-Sourced PyTorch Implementations

[S4] ⦿ `fsc-clip`: Implementations of training CLIP models with hard negative captions, related to [W2], [C6].
  ○ Supports training on 3 image-text datasets with unified evaluation using `vl_compo`.

[S3] ⦿ `vl_compo`: A unified evaluation toolkit for compositional reasoning and multi-modal tasks, related to [W2], [C6].
  ○ Incorporates a wide range of models and benchmarks for evaluation, continuously evolving to stay up-to-date.

[S2] ⦿ `retriever`: An implementation of the second-place solution in the 2023 NICE Challenge, related to [W1].
  ○ Enhances BLIP-2 with retrieval-augmented image captioning for world knowledge and specific details.

[S1] ⦿ `daso`: A unified codebase for semi-supervised learning (SSL), related to [C3].
  ○ Supports training of SSL algorithms, including techniques for class-imbalance.
  ○ Integrates 6 base SSL algorithms and 4 class-imbalanced methods within a single codebase.

Acedemic Services

Conference Reviewer
  ○ Conference on Computer Vision and Pattern Recognition (**CVPR**): 2022 – 2024
  ○ European Conference on Computer Vision (**ECCV**): 2022 (Outstanding), 2024
  ○ International Conference on Computer Vision (**ICCV**): 2023
  ○ Conference on Neural Information Processing Systems (**NeurIPS**): 2023, 2024
  ○ International Conference on Learning Representations (**ICLR**): 2024, 2025
  ○ Winter Conference on Applications of Computer Vision (**WACV**): 2024, 2025
  ○ International Conference on Acoustics, Speech, and Signal Processing (**ICASSP**): 2024, 2025
  ○ AAAI Conference on Artificial Intelligence (**AAAI**): 2025

Journal Reviewer
  ○ International Journal of Computer Vision (**IJCV**): 2023, 2024

Teaching Experience

Teaching Assistant (TA) at EE, KAIST
  ○ EE735: Computer Vision (Fall, 2020)
  ○ EE898: Advanced Topics in Deep Learning for Robotics and Computer Vision (Spring, 2020)
  ○ EE405: Electronics Design Lab (Fall, 2019)