

Sangmin Woo

PH.D. CANDIDATE IN EE @ KAIST

291, Daehak-ro, Yuseong-gu, Daejeon, 34141, Rep. of KOREA

[✉ smwoo95@kaist.ac.kr](mailto:smwoo95@kaist.ac.kr) | [🏠 sangminwoo.github.io](https://sangminwoo.github.io) | [📄 sangminwoo](https://github.com/sangminwoo) | [🎓 Sangmin Woo](https://www.sangminwoo.com)

I am currently pursuing a Ph.D. degree in Electrical Engineering at [KAIST](#). In 2021, I completed an M.S. degree in Electrical Engineering and Computer Science at [GIST](#). Prior to that, I obtained a B.S. degree in Electrical Engineering from [KNU](#) in 2019.

I thrive on creative challenges and enjoy building strong relationships along the way. Explore my academic journey below, and contact me directly to learn more.

Research Interest

Humans are inherently multi-modal learners, with **vision** playing a pivotal role in shaping our understanding of the world. I am passionate about bridging the gap between machine perception and human-level understanding by harnessing the potential of **multi-modal learning**.

My work explores the following, but not limited to:

- **Multi-modal Learning**
 - > High-level: Vision + $X \in \{\text{Language, Audio, Sketch, etc.}\}$
 - > Low-level: RGB + $X \in \{\text{Depth, IR, Flow, etc.}\}$
- **Video / Image Understanding**
- **Generation & Diffusion Models**

Research Experience

Amazon AWS AI

RESEARCH INTERN

Remote

Sep. 2024 - Mar. 2025

Amazon AWS AI

RESEARCH INTERN

Santa Clara, CA, United States

Jun. 2024 - Sep. 2024

Robot Vision Team @ NAVER LABS

RESEARCH INTERN

Suwon, Korea

Apr. 2023 - Aug. 2023

- My primary focus involved pushing the boundaries of **multi-modal multi-task learning**, aiming to tackle a complex challenge: given inputs in the form of RGB imagery, partially captured depth information, and incomplete semantic segmentation, the objective is to create a model that could simultaneously refine the depth perception and complete the missing segments in the semantic segmentation.

Publication

(C: conference, J: journal, P: preprint)

2024

[C15] RITUAL: Random Image Transformations as a Universal Anti-hallucination Lever in LVLMS

2024

SUBMITTED TO ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS)

Multi-modal

[Paper](#) | [Code](#) | [Project](#)

Sangmin Woo*, Jaehyuk Jang*, Donguk Kim*, Changick Kim (*: Equal Contribution)

- [C14] Don't Miss the Forest for the Trees: Attentional Vision Calibration for Large Vision Language Models** 2024
 SUBMITTED TO ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS) Multi-modal
[Paper](#) | [Code](#) | [Project](#)
Sangmin Woo*, Donguk Kim*, Jaehyuk Jang*, Changick Kim (*: Equal Contribution)
- [C13] Diffusion Model Patching via Mixture-of-Prompts** 2024
 SUBMITTED TO ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS) Generation
[Paper](#) | [Code](#) | [Project](#)
 Seokil Ham*, **Sangmin Woo***, Jinyoung Kim, Hyojun Go, Byeongjun Park, Changick Kim (*: Equal Contribution)
- [C12] Flow-Assisted Motion Learning Network for Weakly-Supervised Group Activity Recognition** 2024
 EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV) Multi-modal & Video Understanding
[Paper](#)
 Muhammad Adi Nugroho, **Sangmin Woo**, Sumin Lee, Jinyoung Park, Yooseung Wang, Donguk Kim, Changick Kim
- [C11] Spatio-Temporal Proximity-Aware Dual-Path Model for Panoramic Activity Recognition** 2024
 EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV) Video Understanding
[Paper](#)
 Sumin Lee, Yooseung Wang, **Sangmin Woo**, Changick Kim
- [C10] Switch Diffusion Transformer: Synergizing Denoising Tasks with Sparse Mixture-of-Experts** 2024
 EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV) Generation
[Paper](#) | [Code](#) | [Project](#)
 Byeongjun Park, Hyojun Go, Jinyoung Kim, **Sangmin Woo**, Seokil Ham*, Changick Kim
- [C9] HarmonyView: Harmonizing Consistency and Diversity in One-Image-to-3D** 2024
 IEEE / CVF COMPUTER VISION AND PATTERN RECOGNITION CONFERENCE (CVPR) Generation
[Paper](#) | [Code](#) | [Project](#) | [Demo](#)
Sangmin Woo*, Byeongjun Park*, Hyojun Go, Jinyoung Kim, Changick Kim (*: Equal Contribution)
- [C8] Denoising Task Routing for Diffusion Models** 2024
 INTERNATIONAL CONFERENCE OF LEARNING REPRESENTATION (ICLR) Generation
[Paper](#) | [Code](#) | [Project](#)
 Byeongjun Park*, **Sangmin Woo***, Hyojun Go*, Jinyoung Kim*, Changick Kim (*: Equal Contribution)
- [C7] Sketch-based Video Object Localization** 2024
 IEEE WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV) Multi-modal & Video Understanding
[Paper](#) | [Code](#)
Sangmin Woo, Soyeong Jeon, Jinyoung Park, Minji Son, Sumin Lee, Changick Kim
- 2023
- [C6] AHFu-Net: Align, Hallucinate, and Fuse Network for Missing Multimodal Action Recognition** 2023
 IEEE INTERNATIONAL CONFERENCE ON VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP) (ORAL PRESENTATION) Multi-modal & Video Understanding
 Muhammad Adi Nugroho, **Sangmin Woo**, Sumin Lee, Changick Kim
- [C5] Multi-modal Social Group Activity Recognition in Panoramic Scene** 2023
 IEEE INTERNATIONAL CONFERENCE ON VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP) Multi-modal & Video Understanding
 Donguk Kim, Sumin Lee, **Sangmin Woo**, Jinyoung Park, Muhammad Adi Nugroho, Changick Kim
- [J6] Cross-Modal Alignment and Translation for Missing Modality Action Recognition** 2023
 COMPUTER VISION AND IMAGE UNDERSTANDING (CVIU) Multi-modal & Video Understanding
[Paper](#)
 Yeonju Park, **Sangmin Woo**, Sumin Lee, Muhammad Adi Nugroho, Changick Kim

- [J5] Modality Mixer Exploiting Complementary Information for Multi-modal Action Recognition** 2023
 IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP) – MAJOR REVISION Multi-modal & Video Understanding
[Paper](#)
 Sumin Lee, [Sangmin Woo](#), Yeonju Park, Muhammad Adi Nugroho, Changick Kim
- [C4] Audio-Visual Glance Network for Efficient Video Recognition** 2023
 IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV) Multi-modal & Video Understanding
[Paper](#)
 Muhammad Adi Nugroho, [Sangmin Woo](#), Sumin Lee, Changick Kim
- [C3] Towards Good Practices for Missing Modality Robust Action Recognition** 2023
 AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE (AAAI) (ORAL PRESENTATION) Multi-modal & Video Understanding
[Paper](#) | [Code](#)
[Sangmin Woo](#), Sumin Lee, Yeonju Park, Muhammad Adi Nugroho, Changick Kim
- [C2] Modality Mixer for Multi-modal Action Recognition** 2023
 IEEE WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV) Multi-modal & Video Understanding
[Paper](#)
 Sumin Lee, [Sangmin Woo](#), Yeonju Park, Muhammad Adi Nugroho, Changick Kim
- ~2022
- [P1] Explore-And-Match: Bridging Proposal-Based and Proposal-Free with Transformer for Sentence Grounding in Videos** 2022
 ARXIV Multi-modal & Video Understanding
[Paper](#) | [Code](#)
[Sangmin Woo](#), Jinyoung Park, Inyong Koo, Sumin Lee, Minki Jeong, Changick Kim
- [J4] Tackling the Challenges in Scene Graph Generation with Local-to-Global Interactions** 2022
 IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS) Multi-modal & Image Understanding
[Paper](#) | [Code](#)
[Sangmin Woo](#), Junhyug Noh, Kangil Kim
- [C1] Temporal Flow Mask Attention for Open-Set Long-Tailed Recognition of Wild Animals in Camera-Trap Images** 2022
 IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP) Image Understanding
[Paper](#)
 Jeongsoo Kim, [Sangmin Woo](#), Byeongjun Park, Changick Kim
- [J3] Impact of Sentence Representation Matching in Neural Machine Translation** 2022
 APPLIED SCIENCES General Learning
[Paper](#)
 Heeseung Jung, Kangil Kim, Jong-Hun Shin, Seung-Hoon Na, SangKeun Jung, [Sangmin Woo](#)
- [J2] What and When to Look?: Temporal Span Proposal Network for Video Relation Detection** 2021
 EXPERT SYSTEMS WITH APPLICATIONS (ESWA) – MAJOR REVISION Video Understanding
[Paper](#) | [Code](#)
[Sangmin Woo](#), Junhyug Noh, Kangil Kim
- [J1] Revisiting Dropout: Escaping Pressure for Training Neural Networks with Multiple Costs** 2021
 ELECTRONICS General Learning
[Paper](#) | [Code](#)
[Sangmin Woo](#), Kangil Kim, Junhyug Noh, Jong-Hun Shin, Seung-Hoon Na

DOMESTIC

Light-Weighted Korean Speech Recognition System for Edge Devices

INSTITUTE OF ELECTRONICS AND INFORMATION ENGINEERS (IEIE)

Yooseung Wang, Sangmin Woo, Changick Kim

2023

General Learning

Light-Weighted Korean Speech Recognition System for Edge Devices

INSTITUTE OF ELECTRONICS AND INFORMATION ENGINEERS (IEIE)

Yooseung Wang, Sangmin Woo, Changick Kim

2023

General Learning

On Learning Relations between Objects in Images

KOREA INSTITUTE OF MILITARY SERVICE AND TECHNOLOGY (KIMST)

Sangmin Woo, Changick Kim

2022

Image Understanding

Effective Trash Classification using Attentional Learning

KOREA SOFTWARE CONGRESS (KSC)

[Code](#)

Sangmin Woo, Soon Ki Jung

2018

Image Understanding

Honors & Awards

Oct, 2023	Invited Paper Talk , Center for Applied Research in Artificial Intelligence (CARAI) Workshop
Dec, 2022	Finalist , 29th HumanTech Paper Award @ Samsung Electronics Co., Ltd.
Dec, 2021	Top Award (\$ 10,000) , LG Electronics Robot Contest @ LG Electronics Co., Ltd.
Nov, 2019	Excellence Award (\$ 500) , Creative Space G A.I&IoT Makerthon @ GIST

Patent

Method for group activity recognition using RGB videos and LiDAR data

KR PATENT APPLICATION

Changick Kim, Jinyoung Park, Donguk Kim, Sumin Lee, Muhammad Adi Nugroho, Sangmin Woo, Yooseung Wang

2023

In Progress

Method and Apparatus for Human Activity Recognition using Accelerometer and Gyroscope Sensors

KR PATENT APPLICATION: 10-2022-0094911

Changick Kim, Inyong Koo, Yeonju Park, Minki Jeong, Sumin Lee, Sangmin Woo

2022

Method and Device for Inferring Dynamic Relationship between Objects in Video

KR PATENT APPLICATION: 10-2021-0125704

Sangmin Woo, Kangil Kim

2021

Scene Graph Generation Apparatus

KR PATENT 10-2254-7680000

Sangmin Woo, Kangil Kim

2021

Academic Activity

I serve as a reviewer in the following conferences and journals.

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	2024 ~
European Conference on Computer Vision (ECCV)	2024 ~
Annual Conference on Neural Information Processing Systems (NeurIPS)	2024 ~
International Conference on Learning Representations (ICLR)	2024 ~
AAAI Conference on Artificial Intelligence (AAAI)	2023 ~
International Conference on Artificial Intelligence and Statistics (AISTATS)	2025 ~
IEEE Transactions on Neural Networks and Learning Systems (TNNLS)	
IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	

Education

Korea Advanced Institute of Science and Technology (KAIST)

PH.D. IN ELECTRICAL ENGINEERING

Daejeon, Korea

Aug. 2021 - Present

Gwangju Institute of Science and Technology (GIST)

M.S. IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Gwangju, Korea

Sep. 2019 - Aug. 2021

Kyungpook National University

B.S. IN ELECTRICAL ENGINEERING (MINOR IN COMPUTER SCIENCE AND ENGINEERING)

Daegu, Korea

Mar. 2013 - Aug. 2019

Project

Scene Text Recognition with Visual Contexts

CENTER FOR SECURITY TECHNOLOGY RESEARCH, KAIST

2024.02 present

Multi-modal Group Activity Recognition

CENTER FOR APPLIED RESEARCH IN ARTIFICIAL INTELLIGENCE (CARAI)

2023.02 present

Sketch-based Video Object Localization

CENTER FOR SECURITY TECHNOLOGY RESEARCH, KAIST

2023.02 2023.11

Multi-modal Action Recognition

CENTER FOR APPLIED RESEARCH IN ARTIFICIAL INTELLIGENCE (CARAI)

2021.09 2022.12

Development of Precise Content Identification Technology based on Relationship Analysis for Maritime Vessels/Structure

MINISTRY OF SCIENCE AND ICT (MSIT)

2021.09 2021.12