Sangmin **Woo**

PH.D. CANDIDATE IN EE @ KAIST

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

smwoo95@kaist.ac.kr | sangminwoo.github.io | sangminwoo | Sangmin Woo

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 AI**
  - > High-level: Vision + X ∈ {Language, Audio, Sketch, *etc.*}
  - > Low-level: RGB + X ∈ {Depth, IR, Flow, *etc.*}
- **Video / Image Understanding**
- **Generation & Diffusion Models**

Research Experience

Amazon AWS AI — *Remote*

RESEARCH INTERN — *Sep. 2024 - Mar. 2025*

Amazon AWS AI — *Santa Clara, CA, United States*

RESEARCH INTERN — *Jun. 2024 - Sep. 2024*

Robot Vision Team @ NAVER LABS — *Suwon, Korea*

RESEARCH INTERN — *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

2025

Black-Box Visual Prompt Engineering for Mitigating Object Hallucination in Large Vision Language Models — *2025*

ARXIV — *Multi-modal*

Paper

Sangmin Woo, Kang Zhou, Yun Zhou, Shuai Wang, Sheng Guan, Haibo Ding, Lin Lee Cheong

Diffusion Model Patching via Mixture-of-Prompts — *2025*

AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE (**AAAI**) — *Generation*

Paper | Code | Project

Seokil Ham*, **Sangmin Woo***, Jinyoung Kim, Hyojun Go, Byeongjun Park, Changick Kim (***: Equal Contribution**)

2024

RITUAL: Random Image Transformations as a Universal Anti-hallucination Lever in LVLMs *2024*

Arxiv *Multi-modal*

Paper | Code | Project

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

Don't Miss the Forest for the Trees: Attentional Vision Calibration for Large Vision Language Models *2024*

Arxiv *Multi-modal*

Paper | Code | Project

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

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

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

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

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**)

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**)

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

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

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

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

Modality Mixer Exploiting Complementary Information for Multi-modal Action Recognition *2023*

Computer Vision and Image Understanding (**CVIU**)) – Major Revision *Multi-modal & Video Understanding*

Paper

Sumin Lee, **Sangmin Woo**, Yeonju Park, Muhammad Adi Nugroho, Changick Kim

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

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

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

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

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

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

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**

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

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 *2023*

Institute of Electronics and Information Engineers (**IEIE**) *General Learning*

Yooseung Wang, **Sangmin Woo**, Changick Kim

Light-Weighted Korean Speech Recognition System for Edge Devices *2023*

Institute of Electronics and Information Engineers (**IEIE**) *General Learning*

Yooseung Wang, **Sangmin Woo**, Changick Kim

On Learning Relations between Objects in Images *2022*

Korea Institute of Military Service and Technology (**KIMST**) *Image Understanding*

Sangmin Woo, Changick Kim

Effective Trash Classification using Attentional Learning *2018*

Korea Software Congress (**KSC**) *Image Understanding*

Code

Sangmin Woo, Soon Ki Jung

Honors & Awards

Dec, 2024	**Finalist,** Qualcomm Innovation Fellowship 2024 Korea
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 *2023*

KR Patent Application *In Progress*

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

Method and Appratus for Human Activity Recognition using Accelerometer and Gyroscope Sensors *2022*

KR Patent Application: 10-2022-0094911

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

Method and Device for Inferring Dynamic Relationship between Objects in Video *2021*

KR Patent Application: 10-2021-0125704

Sangmin Woo, Kangil Kim

Scene Graph Generation Apparatus *2021*

KR Patent 10-2254-7680000

Sangmin Woo, Kangil Kim

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 ~
International Conference on Machine Learning (ICML)	2025 ~
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) *Daejeon, Korea*

Ph.D. in Electrical Engineering *Aug. 2021 - Present*

Gwangju Institute of Science and Technology (GIST) *Gwangju, Korea*

M.S. in Electrical Engineering and Computer Science *Sep. 2019 - Aug. 2021*

Kyungpook National University *Daegu, Korea*

B.S. in Electrical Engineering (Minor in Computer Science and Engineering) *Mar. 2013 - Aug. 2019*

Project

Scene Text Recognition with Visual Contexts *2024.02 2024.12*

Center for Security Technology Research, KAIST

Multi-modal Group Activity Recognition *2023.02 2024.02*

Center for Applied Research in Artificial Intelligence (CARAI)

Sketch-based Video Object Localization *2023.02 2023.11*

Center for Security Technology Research, KAIST

Multi-modal Action Recognition *2021.09 2022.12*

Center for Applied Research in Artificial Intelligence (CARAI)

Development of Precise Content Identification Technology based on Relationship Analysis for Maritime Vessels/Structure *2021.09 2021.12*

Ministry of Science and ICT (MSIT)