

Sungyeon Kim

ksy9597@gmail.com | sung-yeon-kim.github.io

 [sung-yeon-kim](#) |  [sungyeonkim-b47b0a242](#) |  [sungyeonkim-05Ff7hcAAAAJ](#)

EMPLOYMENT

Applied Scientist II

Aug 2025 – Jan 2026

Palo Alto, CA

Amazon, Core Search – Visual Search (Amazon Lens Live)

- Contributed to the research and development of multimodal retrieval and result-verification components for Amazon Lens Live, a real-time video-based visual search system serving millions of users.
- Designed and implemented a client-side query verification module that increased precision and reduced false positives in real-time visual search.
- Built a server-side retrieval validation framework ensuring intent-consistent and trustworthy search results.
- Enhanced the MLLM-based contextual retrieval pipeline to capture reasoning beyond pure visual similarity.

Postdoctoral Researcher

Feb 2025 – Mar 2025

Pohang, South Korea

POSTECH, Computer Vision Lab

- Conducted postdoctoral research under Prof. Suha Kwak on generative retrieval frameworks for human-centric multimodal video search and compact representation learning.
- Designed a generative retrieval architecture bridging textual and visual semantics for scalable multimodal retrieval.

Applied Scientist Intern

Jun 2024 – Sep 2024

Palo Alto, CA

Amazon, Core Search – Visual Search

- Researched and implemented generative retrieval models for universal multimodal search, exploring decoder-based retrieval generation as an alternative to dense embedding retrieval.
- First-author, *GENIUS: A Generative Framework for Universal Multimodal Search*, CVPR 2025.
- Collaborated with Xinliang Zhu, Xiaofan Lin, and Muhammet Bastan under Douglas Gray.

Research Collaborator

Dec 2022 – Sep 2023

Cambridge, MA (Remote)

MIT-IBM Watson AI Lab

- Collaborated with IBM Research on parameter-efficient and unified metric learning for large-scale, cross-domain representation transfer.
- Developed a parameter-efficient transfer learning framework for deep metric learning under distribution shift.
- First-author, *Parameter-efficient Transfer for Unified Distance Metric Learning*, WACV 2025.

Research Intern

Apr 2022 – Jul 2022

Seongnam, S. Korea (Remote)

Naver, Vision Team

- Researched with Geonmo Gu and Byungsoo Ko on adapting NLP self-supervised pretraining (e.g., BERT, ELECTRA) to vision for robust representation learning.
- Implemented visual pretraining frameworks inspired by masked language modeling and evaluated them on large-scale image datasets.

EDUCATION

POSTECH (Pohang University of Science and Technology)

Pohang, South Korea

Ph.D. in Computer Science and Engineering

Sep. 2018 – Feb. 2025

- Advised by Prof. Suha Kwak.
- Dissertation: Towards Retrieval at Scale via Compact Embeddings and Generative Modeling.
- Committee: Profs. Suha Kwak, Minsu Cho, Seungyong Lee, Jungseul Ok, Bohyung Han.
- Research focuses on deep metric learning, multimodal and generative learning, and representation learning.
- Additional research experience includes domain generalization, active learning, and multimodal reasoning with large language and multimodal large language models (LLMs / MLLMs).

DGIST (Daegu Gyeongbuk Institute of Science and Technology)

Daegu, South Korea

B.S. in Undergraduate Studies

Mar. 2014 – Feb. 2018

SELECTED PROJECTS

Harrier: Client–Server Multistage Verification Framework for Amazon Lens Live

- Led the design and development of Harrier, a client–server multistage verification system ensuring retrieval precision and user-intent alignment in Amazon Lens Live.
- Implemented a **client-side query verifier** integrated with real-time object detection to validate selected video frames and objects, prompting re-selection when queries were unreliable.
- Developed a **server-side retrieval verifier** that classifies query–ASIN pairs as exact, similar, or irrelevant with calibrated confidence, filtering incorrect or low-confidence matches.
- Deployed end-to-end in Lens Live pipelines to **filter false positives and trigger MLLM-based re-search** for improved contextual precision.

MLLM-based Contextual Retrieval System

- Built a **multimodal large language model (MLLM)** retrieval pipeline that performs reasoning-based **re-search** when Harrier detects low-confidence or misaligned results.
- Leveraged multimodal reasoning over textual queries, detected objects, and contextual signals to enable **semantic retrieval refinement beyond visual similarity**.
- Integrated with Harrier as a re-search engine to enhance contextual understanding and user satisfaction in Lens Live.

GENIUS: Generative Retrieval for Universal Multimodal Search | CVPR 2025

- Proposed a **generative retrieval framework** replacing dense embedding search with text-conditioned retrieval generation for text, image, and multimodal queries.
- Achieved comparable accuracy to dense retrieval while **dramatically improving retrieval speed and scalability** at large scale.
- First-author publication at CVPR 2025; integrated prototype models into Amazon Core Search’s experimental generative retrieval stack.

Proxy Anchor Loss for Deep Metric Learning | CVPR 2020

- Proposed a novel **proxy-based loss function** that unifies classification and pair-based metric learning objectives via anchor-proxy formulation.
- Achieved state-of-the-art retrieval performance across multiple benchmarks and enabled stable training.
- First-author publication at CVPR 2020 with over **530+ citations**, establishing a widely adopted foundation for metric learning research.

TECHNICAL SKILLS

Programming & Infrastructure: Python, C++, CUDA, PyTorch, TensorFlow, PySpark, AWS (S3, EC2), Docker, Bash

Machine Learning & AI: Multimodal Retrieval, Deep Metric Learning, Self-supervised Learning, Generative Modeling, Large-scale Representation Learning, Vision–Language Models (CLIP, BLIP-2, LLaVA), Large Language Models (LLMs), Multimodal LLMs (MLLMs), Retrieval-Augmented Generation (RAG), Active Learning, Domain Generalization, parameter-efficient learning

Distributed Training & Systems: PyTorch Distributed, DeepSpeed, Accelerate, Data/Model Parallelism, Mixed-precision Optimization, GPU Cluster Orchestration, Large-batch Training Pipelines

Languages: English (Professional), Korean (Native)

PUBLICATIONS

Works in Submission

- [1] Dosung Lee, Boyoung Kim, Minyoung Kim, Junyoung Sung, Sangwon Jung, **Sungyeon Kim**, Paul Hongsuck Seo

Context-Related Visual Question Answering: Toward Unbiased Multimodal Understanding for RAG
In submission, conducted in collaboration with Korea University and KAIST.

International Conference

- [1] **Sungyeon Kim**, Xinliang Zhu, Xiaofan Lin, Muhammet Bastan, Douglas Gray, Suha Kwak
GENIUS: A Generative Framework for Universal Multimodal Search
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2025
- [2] Boseung Jeong, Jicheol Park, **Sungyeon Kim**, Suha Kwak
Learning Audio-guided Video Representation with Gated Attention for Video-Text Retrieval
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2025
(Oral Presentation, 3.3%)
- [3] **Sungyeon Kim**, Donghyun Kim, Suha Kwak
Learning Unified Distance Metric Across Diverse Data Distributions with Parameter-Efficient Transfer Learning
IEEE/CVF Winter Conference on Applications of Computer Vision (**WACV**), 2025
- [4] **Sungyeon Kim**, Boseung Jeong, Donghyun Kim, Suha Kwak
Efficient and Versatile Robust Fine-Tuning of Zero-shot Models
European Conference on Computer Vision (**ECCV**), 2024
- [5] Sohyun Lee, Namyup Kim, **Sungyeon Kim**, Suha Kwak
FREST: Improving Robustness of Semantic Segmentation via Source-free Domain Adaptation with Feature Restoration
European Conference on Computer Vision (**ECCV**), 2024
- [6] Junhyeong Cho, Gilhyun Nam, **Sungyeon Kim**, Hunmin Yang, Suha Kwak
PromptStyler: Prompt-driven Style Generation for Source-free Domain Generalization
IEEE/CVF International Conference on Computer Vision (**ICCV**), 2023
- [7] **Sungyeon Kim**, Boseung Jeong, Suha Kwak
HIER: Metric Learning Beyond Class Labels via Hierarchical Regularization
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2023
- [8] Kyungmoon Lee, **Sungyeon Kim**, Suha Kwak
Cross-Domain Ensemble Distillation for Domain Generalization
European Conference on Computer Vision (**ECCV**), 2022
- [9] Sehyun Hwang, Sohyun Lee, **Sungyeon Kim**, Jungseul Ok, Suha Kwak
Combating Label Distribution Shift for Active Domain Adaptation
European Conference on Computer Vision (**ECCV**), 2022
- [10] **Sungyeon Kim**, Dongwon Kim, Minsu Cho, Suha Kwak
Self-Taught Metric Learning without Labels
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022
- [11] Kyungmoon Lee, **Sungyeon Kim**, Seunghoon Hong, Suha Kwak
Learning to Generate Novel Classes for Deep Metric Learning for Improved Metric Learning
British Machine Vision Conference (**BMVC**), 2021
- [12] **Sungyeon Kim**, Dongwon Kim, Minsu Cho, Suha Kwak
Embedding Transfer with Label Relaxation for Improved Metric Learning
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2021
- [13] **Sungyeon Kim**, Dongwon Kim, Minsu Cho, Suha Kwak
Proxy Anchor Loss for Deep Metric Learning
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2020

- [14] **Sungyeon Kim**, Minkyo Seo, Ivan Laptev, Minsu Cho, Suha Kwak
Deep Metric Learning Beyond Binary Supervision
IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2019
(Oral Presentation, 5.58%)

HONORS & AWARDS

- **Winner**, Alumni Award, POSTECH, 2025
- **Winner**, Qualcomm Innovation Fellowship Korea, Qualcomm Technologies Inc., 2024
- **Winner**, Google PhD Fellowship Program, Google LLC, 2023
- **Winner**, BK21 Best Paper Award, Dept. CSE, POSTECH, 2023
- **Winner**, Qualcomm Innovation Fellowship Korea, Qualcomm Technologies Inc., 2022
- **Winner**, BK21 Best Paper Award, Dept. CSE, POSTECH, 2022
- **Gold Prize**, IPIU Best Paper Award, Workshop on Image Processing and Image Understanding (IPIU), 2022
- **Outstanding Reviewer**, CVPR, IEEE, 2022
- **2nd Place**, ICT Paper Contest, Etnews, Webcash Group, and KSFC, 2021
- **Winner**, SKT AI Fellowship, SK Telecom Co., Ltd, 2021
- **Winner**, POSTECHIAN Fellowship, POSTECH, 2021
- **Grand Prize**, IPIU Best Paper Award, Workshop on Image Processing and Image Understanding (IPIU), 2021
- **Winner**, Naver Ph.D Fellowship, NAVER Corp., 2020
- **Winner**, Qualcomm Innovation Fellowship Korea, Qualcomm Technologies Inc., 2020

ACADEMIC SERVICE

- Have been served as a reviewer for international conferences, such as **CVPR**, **ICCV**, **ECCV**, **ICLR**, **ICML**, **NeurIPS**, **AAAI**, **BMVC**, **ACCV**, **WACV**, and so on.
- Have been served as a reviewer for international journals, such as **TPAMI**, **IJCV**, and **TIP**.

TALKS

- GENIUS: A Generative Framework for Universal Multimodal Search, **Google**, Mountain View, CA, 2025
- Towards Retrieval at Scale via Compact Embeddings and Generative Modeling, **CVPR Doctoral Consortium**, Nashville, TN, 2025
- Efficient and Versatile Robust Fine-Tuning of Zero-shot Models, **Amazon**, Palo Alto, CA, 2025
- Transcending Binary Supervision for Improved Metric Learning, **Artificial Intelligence Graduate School (AIGS) Symposium**, Pohang, Republic of Korea, 2023
- Hierarchical Regularization for Metric Learning Applications, Qualcomm Innovation Fellowship Korea, **Qualcomm**, Seoul, Republic of Korea, 2022
- Efficient Label Relaxation Techniques for Deep Metric Learning, Qualcomm Innovation Fellowship Korea, **Qualcomm**, Seoul, Republic of Korea, 2020

- Structured and Continuous Labels for Deep Metric Learning, **Korea Computer Congress**, Jeju, Republic of Korea, 2019
- Implementing Triplet Loss and Contrastive Loss in Metric Learning, **Samsung Advanced Institute of Technology**, Suwon, Republic of Korea, 2019
- Metric Learning: From Distance Metric Learning to Deep Metric Learning, **Samsung Advanced Technology Training Institute**, Suwon, Republic of Korea, 2018
- C Programming Tutorial for Beginners, **Daegu Software High School**, Daegu, Republic of Korea, 2017

PATENTS

- Rehabilitation program creation method for muscle treatment and rehabilitation program providing apparatus for performing the method, KR101648638B1, Republic of Korea