

Jiyoung Lee

RESEARCH SCIENTIST · NAVER AI LAB

✉ lee.j@navercorp.com | 🏠 <https://lee-jiyoung.github.io/>

Research Interest

Multimodal learning, computer vision, machine learning

Education

Yonsei University

PH.D. IN SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

- Dissertation: Comprehensive human-centric video understanding: learning spatio-temporal and audio-visual attention
- Supervised by Prof. Kwanghoon Sohn.

Seoul, S.Korea

Mar. 2016 - Feb. 2022

Yonsei University

B.S. IN SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

Seoul, S.Korea

Mar. 2012 - Feb. 2016

Experience

NAVER AI Lab

RESEARCH SCIENTIST

- ML Research

Seongnam, S. Korea

Dec. 2021 -

Adobe Research

INTERN

- Creative Intelligence Lab
- Remote working with Dr. Justin Salamon and Dr. Dingzeyu Li
- Research topic: Audio-visual Few-shot Event Detection

San Francisco, USA

May. 2021 - Nov. 2021

Microsoft Research

STUDENT CO-WORKER

- Human Understanding and Empathy Group, and Computer Vision Group.
- Remote working with Dr. Daniel McDuff, Dr. Yale Song, and Dr. Vibhav Vineet.
- Research topic: Causal Discovery and Simulation for Autonomous Driving.

Redmond, USA

Apr. 2020 - Feb. 2021

Publication

* indicates equal contribution, underline indicates corresponding author(s)

Selected publications.....

“Imaginary Voice: Face-styled Diffusion Model for Text-to-Speech”

JIYOUNG LEE, JOON SON CHUNG, AND SOO-WHAN CHUNG

- IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Jun. 2023

“Looking into Your Speech: Learning Cross-modal Affinity for Audio-visual Speech Separation”

JIYOUNG LEE*, SOO-WHAN CHUNG*, SUNOK KIM, HONG-GOO KANG, AND KWANGHOON SOHN

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Jun. 2021

“Context-Aware Emotion Recognition Networks”

JIYOUNG LEE, SEUNGRYONG KIM, SUNOK KIM, JUNGIN PARK, AND KWANGHOON SOHN

- IEEE/CVF International Conference on Computer Vision (ICCV)

Oct. 2019

All publications.....

“Dense Text-to-Image Generation with Attention Modulation”

YUNJI KIM, JIYOUNG LEE, JIN-HWA KIM, JUNG-WOO HA, AND JUN-YAN ZHU

- IEEE/CVF International Conference on Computer Vision(ICCV)

Oct. 2023

- “Hierarchical Visual Primitive Experts for Compositional Zero-Shot Learning”**
 HANJAE KIM, **JIYOUNG LEE**, SEONGHEON PARK, AND KWANGHOON SOHN Oct. 2023
 • IEEE/CVF International Conference on Computer Vision (ICCV)
- “Panoramic Image-to-Image Translation”**
 SOOHYUN KIM, JUNHO KIM, TAEKYUNG KIM, HWAN HEO, SEUNGRYONG KIM, **JIYOUNG LEE**, AND JIN-HWA KIM Apr. 2023
 • arXiv preprint arXiv:2304.04960, <https://arxiv.org/abs/2304.04960>
- “Let 2D Diffusion Model Know 3D-Consistency for Robust Text-to-3D Generation”**
 JUNYOUNG SEO, WOOSSEOK JANG, MIN-SEOP KWAK, JAEHOON KO, HYEONSU KIM, JUNHO KIM, JIN-HWA KIM, **JIYOUNG LEE**, AND SEUNGRYONG KIM Mar. 2023
 • arXiv preprint arXiv:2303.07937, <https://arxiv.org/abs/2303.07937>
- “Robust Camera Pose Refinement for Multi-Resolution Hash Encoding”**
 HWAN HEO, TAEKYUNG KIM, **JIYOUNG LEE**, JAEWON LEE, SOOHYUN KIM, HYUNWOO J KIM, AND JIN-HWA KIM Jul. 2023
 • International Conference on Machine Learning (ICML).
- “Semi-Parametric Video-Grounded Text Generation”**
 SUNGDONG KIM, JIN-HWA KIM, **JIYOUNG LEE**, AND MINJOON SEO Jan. 2023
 • arXiv preprint arXiv:2301.11507, <https://arxiv.org/abs/2301.11507>
- “Three Recipes for Better 3D Pseudo-GTs of 3D Human Mesh Estimation in the Wild”**
 GYEONGSIK MOON, HONGSUK CHOI, SANGHYUK CHUN, **JIYOUNG LEE**, AND SANGDOO YUN Jun. 2023
 • IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (CVPRW).
- “Dual-path Adaptation from Image to Video Transformers”**
 JUNGIN PARK*, **JIYOUNG LEE***, AND KWANGHOON SOHN Jun. 2023
 • IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- “Imaginary Voice: Face-styled Diffusion Model for Text-to-Speech”**
JIYOUNG LEE, JOON SON CHUNG, AND SOO-WHAN CHUNG Jun. 2023
 • IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- “MIDMs: Matching Interleaved Diffusion Models for Exemplar-based Image Translation”**
 JUNYOUNG SEO, GYUSEONG LEE, SEOKJU CHO, **JIYOUNG LEE**, AND SEUNGRYONG KIM Feb. 2023
 • Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI).
- “Language-free Training for Zero-shot Video Grounding”**
 DAHYE KIM, JUNGIN PARK, **JIYOUNG LEE**, SEONGHEON PARK, AND KWANGHOON SOHN Jan. 2023
 • IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).
- “Language-Guided Recursive Spatiotemporal Graph Modeling for Video Summarization”**
 JUNGIN PARK, **JIYOUNG LEE**, AND KWANGHOON SOHN Dec. 2022
 • IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). (Under Review)
- “Mutual Information Divergence: A Unified Metric for Multimodal Generative Models”**
JIN-HWA KIM, YUNJI KIM, **JIYOUNG LEE**, KANG MIN YOO, AND SANG-WOO LEE Nov. 2022
 • Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS).
- “PointFix: Learning to Fix Domain Bias for Robust Online Stereo Adaptation”**
 KWONYOUNG KIM, JUNGIN PARK, **JIYOUNG LEE**, DONGBO MIN, AND KWANGHOON SOHN Oct. 2022
 • European Conference on Computer Vision (ECCV).
- “Pin the Memory: Learning to Generalize Semantic Segmentation”**
 JIN KIM, **JIYOUNG LEE**, JUNGIN PARK, DONGBO MIN, AND KWANGHOON SOHN Jun. 2022
 • IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- “Probabilistic Representations for Video Contrastive Learning”**
 JUNGIN PARK, **JIYOUNG LEE**, IG-JAE KIM, AND KWANGHOON SOHN Jun. 2022
 • IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- “CausalCity: Complex Simulations with Agency for Causal Discovery and Reasoning”**
 DANIEL McDUFF, YALE SONG, **JIYOUNG LEE**, VIBHAV VINEET, SAI VEMPRALA, HADI SALMAN, SHUANG MA, KWANGHOON SOHN, AND ASHISH KAPOOR Jun. 2022
 • Conference on Causal Learning and Reasoning (CLear).

“Multi-domain Unsupervised Image-to-Image Translation with Appearance Adaptive Convolution”

SOMI JEONG, **JIYOUNG LEE**, AND KWANGHOON SOHN

Jun. 2022

- IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

“Wide and Narrow: Video Prediction from Context and Motion”

JAEHOON CHO, **JIYOUNG LEE**, CHANGJAE OH, WONIL SONG, AND KWANGHOON SOHN

Nov. 2021

- British Machine Vision Conference (BMVC).

“Looking into Your Speech: Learning Cross-modal Affinity for Audio-visual Speech Separation”

JIYOUNG LEE*, SOO-WHAN CHUNG*, SUNOK KIM, HONG-GOO KANG, AND KWANGHOON SOHN

Jun. 2021

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

“Bridge to Answer: Structure-aware Graph Interaction Network for Video Question Answering”

JUNGIN PARK, **JIYOUNG LEE**, AND KWANGHOON SOHN

Jun. 2021

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

“Learning Discriminative Action Tubelets for Weakly-supervised Action Detection”

JIYOUNG LEE, SEUNGRYONG KIM, SUNOK KIM, AND KWANGHOON SOHN

May. 2021

- Pattern Recognition (PR). (Under Review)

“Self-balanced Learning for Domain Generalization”

JIN KIM, **JIYOUNG LEE**, JUNGIN PARK, DONGBO MIN, AND KWANGHOON SOHN

Sep. 2021

- IEEE International Conference on Image Processing (ICIP).

“SumGraph: Video Summarization via Recursive Graph Modeling”

JUNGIN PARK*, **JIYOUNG LEE***, IG-JAE KIM, AND KWANGHOON SOHN

Aug. 2020

- European Conference on Computer Vision (ECCV)

“Multi-modal Recurrent Attention Networks for Facial Expression Recognition”

JIYOUNG LEE, SUNOK KIM, SEUNGRYONG KIM, AND KWANGHOON SOHN

May. 2020

- IEEE Transaction on Image Processing (TIP). vol. 29, pp. 6977–6991 (Impact Factor: **11.041**)

“Context-Aware Emotion Recognition Networks”

JIYOUNG LEE, SEUNGRYONG KIM, SUNOK KIM, JUNGIN PARK, AND KWANGHOON SOHN

Oct. 2019

- IEEE/CVF International Conference on Computer Vision (ICCV)

“Video Summarization by Learning Relationships between Action and Scene”

JUNGIN PARK, **JIYOUNG LEE**, SANGRYUL JEON, AND KWANGHOON SOHN

Oct. 2019

- IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)

“Graph Regularization Network with Semantic Affinity for Weakly-supervised Temporal Action Localization”

JUNGIN PARK, **JIYOUNG LEE**, SANGRYUL JEON, SEUNGRYONG KIM, AND KWANGHOON SOHN

Sep. 2019

- IEEE International Conference on Image Processing (ICIP)

“Audio-Visual Attention Networks for Emotion Recognition”

JIYOUNG LEE, SUNOK KIM, SEUNGRYONG KIM, AND KWANGHOON SOHN

Oct. 2018

- ACM Multimedia Workshop- Workshop on Audio-Visual Scene Understanding for Immersive Multimedia (MMW)

“Learning to Detect, Associate, and Recognize Human Actions and Surrounding Scenes in Untrimmed Videos”

JUNGIN PARK, SANGRYUL JEON, SEUNGRYONG KIM, **JIYOUNG LEE**, SUNOK KIM, AND KWANGHOON SOHN

Oct. 2018

- ACM Multimedia Workshop- The 1st Workshop and Challenge on Comprehensive Video Understanding in the Wild (MMW)

“Spatiotemporal Attention Based Deep Neural Networks for Emotion Recognition”

JIYOUNG LEE, SUNOK KIM, SEUNGRYONG KIM, AND KWANGHOON SOHN

Apr. 2018

- IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

“Automatic 2D-to-3D Conversion using Multi-scale Deep Neural Network”

JIYOUNG LEE, HYUNGJOO JUNG, YOUNGJUNG KIM, AND KWANGHOON SOHN

Sep. 2017

- IEEE International Conference on Image Processing (ICIP)

Patent

“Creating Audio-visual Triggers using Few-shot Learning”

JYOUNG LEE, DINGZEYU LI, AND JUSTIN SALAMON

Aug. 2021

- US patent, under review

“Emotion recognition apparatus and method based on multimodal fusion”

JYOUNG LEE, AND KWANGHOON SOHN

Dec. 2020

- Korean patent, 10-2020-0180996

“Audio-Video Matching Area Detection Apparatus and Method”

JYOUNG LEE, AND KWANGHOON SOHN

Jul. 2019

- Korean patent, 10-2019-0090937

“Apparatus and Method for Recognizing Activity and Detecting Activity Area in Video”

JYOUNG LEE, AND KWANGHOON SOHN

Mar. 2019

- Korean patent, 10-2019-0034501

“Emotion Recognition Apparatus and Method Based on Spatiotemporal Attention”

JYOUNG LEE, AND KWANGHOON SOHN

May. 2018

- Korean patent, 10-2018-0053306

Project Experiences

Development of Multi-modal Data Fusion and Artificial Social Intelligence for Comprehensive Scene Understanding and Forecasting

S.Korea

FUNDED BY MINISTRY OF SCIENCE, SOUTH KOREA

Mar. 2021 – Dec. 2021

- Developed artificial social intelligence based on scene recognition and reasoning for future forecasting.

To create AI systems that act appropriately and effectively in novel situations that occur in open worlds

S.Korea

FUNDED BY INSTITUTE OF INFORMATION & COMMUNICATION TECHNOLOGY, SOUTH KOREA

Mar. 2020 – Dec. 2021

- Developed algorithms for autonomous delivery robot that can robust perform computer vision tasks in real-world environments.
- Developed an algorithm for domain generalization using meta-learning.

Fundamental Study of Vision Algorithms for Comprehensive and Through Understanding of Videos

S.Korea

FUNDED BY MINISTRY OF SCIENCE, ICT AND FUTURE PLANNING, SOUTH KOREA.

Sep. 2017 – Dec. 2020

- Developed algorithms for scene understanding and reasoning tasks that can robust perform in real-world videos.
- Construction two video datasets related to emotion and scene recognition.

Intelligent Virtual Reality: Deep Audio-Visual Representation Learning for Multimedia Perception and Reproduction

S.Korea

FUNDED BY INSTITUTE OF INFORMATION & COMMUNICATION TECHNOLOGY, SOUTH KOREA.

Sep. 2017 – Aug. 2019

- Developed an emotion recognition algorithm using audio-visual data.

Emotional Intelligence Technology to Infer Human Emotion and Carry on Dialogue Accordingly

S.Korea

FUNDED BY INSTITUTE OF INFORMATION & COMMUNICATION TECHNOLOGY, SOUTH KOREA.

Sep. 2017 – Jun. 2018

- Implemented a multi-modal dataset using color, depth, and FIR sensors.
- Developed an algorithm for inferring human emotion from multi-spectral sensors.

High Quality 2D-to-Multiview Contents Generation from Large-Scale RGB+D Database

S.Korea

FUNDED BY INSTITUTE OF INFORMATION & COMMUNICATION TECHNOLOGY, SOUTH KOREA.

Mar. 2016 – Aug. 2017

- Implemented depth and stereo data acquisition system using ZED and Kinect v2 cameras.
- Implemented a large scale RGB+depth dataset including indoor and outdoor scenes.
- Developed an algorithm for synthesizing 3D view from single view image (2D-to-3D conversion).

Yonsei University, Dept. of Electrical and Electronic Engineering

S.Korea

TEACHING ASSISTANT.

Mar. 2016 – Jun. 2021

- Digital signal processing, Electrical and electronic engineering experiments: fundamentals, Deep learning Lab.

Professional Activities

Lecture

Seoul National University

TOPICS IN ARTIFICIAL INTELLIGENCE: MULTIMODAL DEEP LEARNING THEORIES AND APPLICATIONS

Fall 2023

Reviewer

NEURIPS 2023, CVPR 2022-2023, ECCV 2022, ICCV 2023, ICML 2023, ICASSP 2023, IEEE TPAMI, IEEE TIP, IEEE Access

Media Coverage

“CausalCity: Introducing a high-fidelity simulation with agency for advancing causal reasoning in machine learning.”

MICROSOFT BLOG, [LINK](#)

Jun. 2021

“A deep learning technique for context-aware emotion recognition.”

TECHXPLORE, [LINK](#)

Aug. 2019

Talks

“Beyond the watching: multimodal scene understanding”

KOREA UNIVERSITY, KOREA AEROSPACE UNIVERSITY

Dec. 2021

“Comprehensive video understanding: from recognition to reasoning.”

MICROSOFT RESEARCH AI BREAKTHROUGHS

Sep. 2020

Honors & Awards

2019	Outstanding 100 National Research Projects , Research Assistant	<i>Ministry of Science and ICT, S. Korea</i>
2019	3rd Award , CoVieW 2019 (IEEE ICCV Challenge)	<i>CoVieW'19</i>
2016	Finalist & Award , University Startup 300	<i>Ministry of Education, S.Korea</i>
2015	Silver Prize , Yonsei Creative Design Challenge	<i>Yonsei University</i>
2015	Award , Campus Reboot Startup Camp	<i>Ministry of Education, S.Korea</i>