Youngjoon Jang

Postdoctoral Researcher, VGG, University of Oxford.

Email: youngjoon@robots.ox.ac.uk
Webpage: https://art-jang.github.io
Google Scholar: link

Research Interest

- My research lies in multimodal AI, aiming to develop models that can effectively understand and generate across audio, visual, and textual modalities. My work spans multimodal large language models, and multimodal generation, as well as sign language recognition and translation.

Education

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, South Korea
Ph.D. in Division of Future Vehicle, School of Electrical Engineering; Advisor: Joon Son Chung — Sep. 2022 – Aug. 2025

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, South Korea
M.S. in Division of Future Vehicle, School of Electrical Engineering; Advisor: In So Kweon — Mar. 2020 – Feb. 2022
- Thesis: Learning Methodology According to Characteristics of Continuous Sign Language Recognition Dataset

Kwangwoon University — Seoul, South Korea
B.S. in Division of Robotics; GPA: 4.3/4.5 — Mar. 2014 – Feb. 2020

Work Experience

University of Oxford — Oxford, UK
Postdoctoral Researcher, VGG. — Nov. 2025 - current

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, South Korea
Postdoctoral Researcher, Multimodal AI Lab. — Sep. 2025 - current

University of Oxford — Oxford, UK
Visiting student, VGG. — May. 2024 - Nov. 2024

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, South Korea
Researcher, Multimodal AI Lab. — Mar. 2022 - Aug. 2022

Publication

International Journals

- [J2] EDNet: A Distortion-Agnostic Speech Enhancement Framework with Gating Mamba Mechanism and Phase Shift-Invariant Training.
  *Doyeop Kwak, **Youngjoon Jang**, Seongyu Kim, and Joon Son Chung*
  - IEEE Transactions on Audio, Speech and Language Processing (TASLP), 2026.
- [J1] Deep Understanding of Sign Language for Sign to Subtitle Alignment.
  Youngjoon Jang, Jeongsoo Choi*, Junseok Ahn, and Joon Son Chung*
  - IEEE Transactions on Multimedia (TMM), 2025.

International Conferences

- [C24] Lost in Translation, Found in Embeddings: Sign Language Translation and Alignment.
  ***Youngjoon Jang**, Lilane Momeni, Zifan Jiang, Joon Son Chung, Gül Varol, and Andrew Zisserman*
  - The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS), 2026.
- [C23] On the Nature of Attention Sink that Shapes Decoding Strategy in MLLMs.
  Suho Yoo, **Youngjoon Jang***, and Joon Son Chung*
  - The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS), 2026.
- [C22] Who Says What: Symbolic Trimodal Binding Mechanisms in Audio-Visual LLMs.
  *Jihoo Jung, **Youngjoon Jang**, and Joon Son Chung*
  - The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS), 2026.
- [C21] Fork-Merge Decoding: Enhancing Multimodal Understanding in Audio-Visual Large Language Models.
  Chaeyoung Jung, **Youngjoon Jang***, Jongmin Choi, and Joon Son Chung*
  - The 2026 Conference on Empirical Methods in Natural Language Processing (EMNLP Findings), 2026.
- [C20] SignMatch: Matching Dictionary Signs to Continuous Sign Language Video.
  *Ryan Wong, **Youngjoon Jang**, Lilane Momeni, Gül Varol, and Andrew Zisserman*
  - The British Machine Vision Association (BMVC), 2026.

- [C19] Segment, Embed, and Align: A Universal Recipe for Aligning Subtitles to Signing.
  *Zifan Jiang, **Youngjoon Jang**, Lilane Momeni, Gül Varol, Sarah Ebling, and Andrew Zisserman*
  - The 64th Annual Meeting of the Association for Computational Linguistics (ACL), 2026.
- [C18] LP-CFM: Perceptual Invariance-Aware Conditional Flow Matching for Speech Modeling.
  *Doyeop Kwak, **Youngjoon Jang**, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2026.
- [C17] FastAV: Efficient Token Pruning for Audio-Visual Large Language Model Inference.
  *Chaeyoung Jung, **Youngjoon Jang**, Seungwoo Lee, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2026.
- [C16] AVCD: Mitigating Hallucinations in Audio-Visual Large Language Models through Contrastive Decoding.
  Chaeyoung Jung, **Youngjoon Jang***, and Joon Son Chung*
  - Neural Information Processing Systems (NeurIPS), 2025.
- [C15] Test-Time Augmentation for Pose-invariant Face Recognition.
  Jaemin Jung, **Youngjoon Jang***, and Joon Son Chung*
  - International Conference on Automatic Face and Gesture Recognition (FG), 2025.
- [C14] Lost in Translation, Found in Context: Sign Language Translation with Contextual Cues.
  Youngjoon Jang, Haran Raajesh*, Lilane Momeni, Gül Varol, Andrew Zisserman*
  - Computer Vision and Pattern Recognition Conference (CVPR), 2025.
- [C13] VoiceDiT: Dual-Condition Diffusion Transformer for Environment-Aware Speech Synthesis.
  Jaemin Jung, Junseok Ahn*, Chaeyoung Jung, Tan Dat Nguyen, **Youngjoon Jang**, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2025.
- [C12] Let Me Finish My Sentence: Video Temporal Grounding with Holistic Text Understanding.
  *Jongbhin Woo, Hyeonggon Ryu, **Youngjoon Jang**, Jae Won Cho, and Joon Son Chung*
  - ACM International Conference on Multimedia (ACMMM), 2024.
- [C11]Faces that Speak: Jointly Synthesising Talking Face and Speech from Text.
  Youngjoon Jang, Ji-Hoon Kim*, Junseok Ahn, Doyeop Kwak, Hongsun Yang, Yooncheol Ju, ILHWAN KIM, Byeong-Yeol Kim, and Joon Son Chung*
  - Computer Vision and Pattern Recognition Conference (CVPR), 2024.
- [C10] Slowfast Network for Continuous Sign Language Recognition.
  Junseok Ahn, **Youngjoon Jang***, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.
- [C9] VoxMM: Rich Transcription of Conversations in the Wild.
  *Doyeop Kwak, Jaemin Jung, Kihyun Nam, **Youngjoon Jang**, Jee-weon Jung, Shinji Watanabe, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.
- [C8] FreGrad: Lightweight and Fast Frequency-aware Diffusion Vocoder.
  *Tan Dat Nguyen, Ji-Hoon Kim, **Youngjoon Jang**, Jaehun Kim, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.
- [C7] TalkNCE: Improving Active Speaker Detection with Talk-aware Contrastive Learning.
  *Chaeyoung Jung, Suyeon Lee, Kihyun Nam, Kyeongha Rho, You Jin Kim, **Youngjoon Jang**, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.
- [C6] Seeing Through the Conversation: Audio-visual Speech Separation based on Diffusion Model.
  *Suyeon Lee, Chaeyoung Jung, **Youngjoon Jang**, Jaehun Kim, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2024.
- [C5] That's What I said: Fully-Controllable Talking Face Generation.
  Youngjoon Jang, Kyeongha Rho*, Jongbhin Woo, Hyeongkeun Lee, Jihwan Park, Youshin Lim, Byeong-Yeol, Kim, and Joon Son Chung*
  - ACM International Conference on Multimedia (ACMMM), 2023.
- [C4] Self-Sufficient Framework for Continuous Sign Language Recognition.
  ***Youngjoon Jang**, Youngtaek Oh, Jae Won Cho, Myungchul Kim, Dong-Jin Kim, In So Kweon, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2023. (Top 3% Paper Recognition)
- [C3] Metric Learning for User-Defined Keyword Spotting.
  Jaemin Jung, Youkyum Kim*, Jihwan Park, Youshin Lim, Byeong-Yeol Kim, **Youngjoon Jang**, and Joon Son Chung*
  - International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2023.
- [C2] 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, and In So Kweon*
  - British Machine Vision Conference (BMVC), 2022.

- [C1] 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, and In So Kweon*
  - International Conference on Automatic Face and Gesture Recognition (FG), 2021.

Awards & Honors

International Competitions

- *1st Place*, R-BIZ Challenge TURTLEBOT3 AUTORACE, Nov. 2018
- *5th Place*, Sumo Robot, International Robot Contest (IRC), Oct. 2018
- *3rd Place*, R-BIZ Challenge TURTLEBOT3 AUTORACE, Sep. 2017
- *5th Place*, RoboCup Iran Open Rescue, Apr. 2017

National Competitions

- *2nd Place*, Science and Technology Specialized University Startup Competition (GIST), Nov. 2021
- *2nd Place*, App Startup Support Program Contest (KAIST), Apr. 2021
- *1st Place*, RoboCup Korea Open Rescue, Feb. 2017

Academic Service

Workshop Organization
* BMVA Symposium on AI for Sign Language Translation, Production, and Linguistics (link)

Teaching

Teaching Assistance (TA) at FV, KAIST
* PD513: Future Vehicle Capstone Design (Spring, 2023)
* PD513: Future Vehicle Capstone Design (Fall, 2022)
* PD806: Automobile Special Topics in Mechanical Engineering (Fall, 2021)

References

Prof. Joon Son Chung KAIST
PhD Supervisor, Multimodal AI Lab
Email: joonsc@kaist.ac.kr

Prof. Andrew Zisserman University of Oxford
Postdoctoral Supervisor, Visual Geometry Group (VGG)
Email: az@robots.ox.ac.uk

Prof. In So Kweon KAIST
MS Supervisor, Robotics and Computer Vision Lab
Email: iskweon@kaist.ac.kr