

Sang-gil Lee

tkdr1f9202@gmail.com • <https://L0SG.github.io> • <https://github.com/L0SG> • <https://www.linkedin.com/in/sang-gil-lee>

I am a research scientist at NVIDIA.

I work on deep generative models for sequences, with a particular focus on speech and audio.

RESEARCH INTERESTS

My research interest spans a wide range of deep generative models (AR, flow, GAN, diffusion, etc.) applied to sequential data. Specifically, I am working on building multi-modal large language models with a focus on audio. During my Ph.D., I focused on time-domain waveform data (speech and audio) to advance generative modeling for audio. I am also broadly interested in speech and audio applications, including text-to-speech, voice conversion, music generation, neural audio codecs, and audio language models.

- **Keywords:** Deep generative models, speech & audio synthesis, sequence & language models

EXPERIENCE

- **NVIDIA**, Applied Deep Learning Research Scientist Jan 2024 – Current
 - I am working on building multi-modal large language models with a focus on audio.
- **Qualcomm AI Research**, Senior Machine Learning Research Engineer Feb 2023 – Jan 2024
 - I developed a framework for Text-to-Speech (TTS) research and development, optimized for deployment on edge devices.
- **NVIDIA**, AI Application Team Sep 2021 – Jan 2022
 - I am the lead author of BigVGAN (ICLR 2023), the state-of-the-art conditional waveform synthesizer.
- **Microsoft Research Asia**, Machine Learning Group Dec 2020 – Jun 2021
 - I worked on a diffusion-based generative model for speech synthesis, PriorGrad (ICLR 2022).
- **Kakao Corporation**, AI Laboratory Jul 2019 – Sep 2019
 - I worked on improving speech synthesis and voice conversion models.
- **Microsoft Research Asia**, Machine Learning Group Dec 2018 – Feb 2019
 - I worked on the Antigen Map Project by applying sequence models to predict antigens from genes.

EDUCATION

Ph.D., Seoul National University, Seoul, South Korea

- Ph.D. in Electrical & Computer Engineering Sep 2016 – Feb 2023
 - Dissertation: Deep Generative Model for Waveform Synthesis
 - Integrated M.S./Ph.D. Program.
- Dual B.S. in Electrical & Computer Engineering / Applied Biology & Chemistry Mar 2010 – Aug 2016
 - Cum Laude

PUBLICATIONS

CONFERENCES

- E. Casanova, R. Langman, P. Neekhara, S. Hussain, J. Li, S. Ghosh, A. Jukić, and S. Lee, “Low Frame-rate Speech Codec: a Codec Designed for Fast High-quality Speech LLM Training and Inference,” in *ICASSP*, 2025.
- Z. Kong*, S. Lee*, D. Ghosal, N. Majumder, A. Mehrish, R. Valle, S. Poria, and B. Catanzaro, “Improving text-to-audio models with synthetic captions,” in *Interspeech SynData4GenAI*, 2024.
- H. Kim, S. Lee, J. Yeom, C.H. Lee, S. Kim, and S. Yoon, “VoiceTailor: Lightweight Plug-In Adapter for Diffusion-Based Personalized Text-to-Speech,” in *Interspeech*, 2024.
- C. Shin, H. Kim, C. Lee, S. Lee, and S. Yoon, “Edit-A-Video: Single Video Editing with Object-Aware Consistency,” in *ACML*, **Best Paper Award**, 2023.
- S. Lee, W. Ping, B. Ginsburg, B. Catanzaro, and S. Yoon, “BigVGAN: A Universal Neural Vocoder with Large-Scale Training,” in *ICLR*, 2023.
- S. Lee, H. Kim, C. Shin, X. Tan, C. Liu, Q. Meng, T. Qin, W. Chen, S. Yoon, and T. Liu, “PriorGrad: Improving Conditional Denoising Diffusion Models with Data-Dependent Adaptive Prior,” in *ICLR*, 2022.
- S. Lee, S. Kim, and S. Yoon, “NanoFlow: Scalable Normalizing Flows with Sublinear Parameter Complexity,” in *NeurIPS*, Vancouver, Canada, 2020.
- S. Kim, S. Lee, J. Song, J. Kim, and S. Yoon, “FloWaveNet : A Generative Flow for Raw Audio,” in *ICML*, Long Beach, CA, USA, 2019.

- S. Lee, J.S. Bae, H. Kim, J.H. Kim, and S. Yoon, "Liver Lesion Detection from Weakly-labeled Multi-phase CT Volumes with a Grouped Single Shot MultiBox Detector," in **MICCAI**, Granada, Spain, 2018.
- S. Park, S. Lee, H. Nam, and S. Yoon, "An Efficient Method to Boosting Performance of Spiking Neural Network Training," in **NIPS Workshop on Computing with Spikes**, Barcelona, Spain, 2016.
- S. Lee and S. Yoon, "Deep Deterministic Policy Gradients as a Proxy for Semi-supervised Deep Learning of Network Intrusion Detection," in **Korea Computer Congress**, Jeju, Korea, 2017.
- J. Lee, Y. Jeon, B. Na, S. Lee and S. Yoon, "Fine Dust Time Series Anomaly Detection using Transfer Entropy and Network Similarity," in **Korea Computer Congress**, Jeju, Korea, 2017.

JOURNALS

- S. Lee*, E. Kim*, J.S. Bae*, J.H. Kim, and S. Yoon, "Robust End-to-End Focal Liver Lesion Detection using Unregistered Multiphase Computed Tomography Images," *IEEE Transactions on Emerging Topics in Computational Intelligence (IEEE TETCI) (Impact Factor: 8.28)*, 2021.
- T. Kim, J.H. Park, S. Lee, S. Kim, J. Kim, J. Lee, and C. Shin, "Small RNA Transcriptome of Hibiscus Syriacus Provides Insights into the Potential Influence of microRNAs in Flower Development and Terpene Synthesis," *Molecules and Cells (Impact Factor: 5.03)*, vol. 40, no. 8, pp. 587, 2017.

ARXIV

- S. Lee*, Z. Kong*, A. Goel, S. Kim, R. Valle, and B. Catanzaro, "ETTA: Elucidating the Design Space of Text-to-Audio Models," in *arXiv preprint:2412.19351*, 2024.
- D. Lee, J. Yoon, J. Song, S. Lee, and S. Yoon, "One-Shot Learning for Text-to-SQL Generation," in *arXiv preprint:1905.11499*, 2019.
- S. Lee, U. Hwang, S. Min, and S. Yoon, "Polyphonic Music Generation with Sequence Generative Adversarial Networks," in *arXiv preprint:1710.11418*, 2017.

REPOSITORIES

relational-rnn-pytorch ★200+

An implementation of DeepMind's Relational Recurrent Neural Networks (Santoro et al. 2018) in PyTorch. Features a full-fledged word language modeling benchmark of the model vs. LSTM.

FloWaveNet ★400+

A Pytorch implementation of ICML 2019 paper "FloWaveNet: A Generative Flow for Raw Audio". (Kim et al., 2019)

WaveFlow ★100+

A PyTorch implementation of ICML 2020 paper "WaveFlow: A Compact Flow-based Model for Raw Audio". (Ping et al., 2020)

NanoFlow ★50+

PyTorch implementation of the NeurIPS 2020 paper "NanoFlow: Scalable Normalizing Flows with Sublinear Parameter Complexity." (Lee et al., 2020)

seqgan-music

Implementation of a paper "Polyphonic Music Generation with Sequence Generative Adversarial Networks" in TensorFlow. (Lee et al., 2017)

grouped-ssd-pytorch

PyTorch implementation of MICCAI 2018 paper "Liver Lesion Detection from Weakly-labeled Multi-phase CT Volumes with a Grouped Single Shot MultiBox Detector" and IEEE TETCI 2021 paper "Robust End-to-End Focal Liver Lesion Detection using Unregistered Multiphase Computed Tomography Images."

OPEN-SOURCE CONTRIBUTION

NVIDIA NeMo ★12000+

Contributed to an open-source implementation of UnivNet vocoder (Jang et al., 2021).

NVIDIA BigVGAN ★900+

Contains an open-source implementation of our work, BigVGAN, the state-of-the-art conditional waveform synthesizer (Lee et al., ICLR 2023).

Microsoft NeuralSpeech ★1400+

Contains an open-source implementation of our work, PriorGrad, an efficient method for accelerating diffusion models for speech synthesis (Lee et al., ICLR 2022).

INVITED TALKS

"Deep Generative Model for Speech and Audio", Soongsil University, 2023

“Towards Universal Neural Waveform Synthesis”, Naver, 2022
“On Neural Waveform Synthesis”, Supertone, 2022
“Prior Enhancement for Deep Generative Models”, Hyundai AIRS, 2022
“Neural Speech Synthesis: a 2021 Landscape”, NVIDIA, 2021
“RNN Plus Alpha: Is RNN the False Prophet?”, Naver CLOVA, 2018

HONORS

Best Paper Award, ACML, 2023
Student Conference Scholarship, Google, 2022
Graduate Student of the Year, DSAIL, Seoul National University, 2019
Best Paper Award, Hyundai AIR Lab (currently AIRS), 2019
Stars of Tomorrow (Excellent Intern), Microsoft Research Asia, 2019
Cum Laude, Seoul National University, 2016
Academic Performance Scholarship, Seoul National University, 2010 - 2016
Academic Scholarship (fully funded), SBS Foundation, 2010 - 2016

REFERENCES

Professor Sungroh Yoon (Ph.D. Advisor)
Professor of Electrical & Computer Engineering, Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea
sryoon@snu.ac.kr • +82-2 880-1401