

WINTER SCHOOL SPEAKERS

09:30-16:30, October 31

Generative AI Techniques for Signal and Information Processing Applications



Professor

Chin-Hui Lee

Georgia Institute of Technology, USA

Deep Regression for Spectral Mapping with Applications to Speech Enhancement, Source Separation and Speech Dereverberation

In this lecture, we present a deep regression framework for solving classical signal processing and spectral mapping problems. Based on Komogorov's Representation Theorem (1957), a multivariate scalar function can be expressed exactly as a superposition of a finite number of inner functions embedded within another linear combination of inner functions. Cybenko (1989) developed a universal approximation theorem showing such a scalar function can be approximated by a superposition of sigmoid functions, inspiring a new wave of neural network algorithms. Barron (1993) later proved that the error in this universal approximation can be tightly bounded and related to the representation power of neural networks in learning theory. In order to make the mapping function learnable and computable for some practical applications, we cast the classical approximation problems into a nonlinear vector-to-vector regression setting using deep neural networks (DNNs) as mapping functions, such that the DNN parameters can be estimated with deep learning and big data configurations.

Many classical speech processing problems, such as enhancement, source separation and dereverberation, can be formulated as finding mapping functions to transform noisy input to clean output spectra. As a result, DNN-transformed speech usually exhibits a good quality and a clear intelligibility under adverse acoustic conditions. Finally, our proposed deep regression framework was also tested on recent challenging tasks in CHiME-2, CHiME-4, CHiME-5, CHiME-6, REVERB and DIHARD evaluations. Based on the top quality achieved in microphone-array based enhancement, separation and dereverberation, our teams often scored the lowest error rates in almost all the above-mentioned open evaluation scenarios.

BIOGRAPHY

Chin-Hui Lee is a professor at School of Electrical and Computer Engineering, Georgia Institute of Technology. Before joining academia in 2001, he had accumulated 20 years of industrial experience ending in Bell Laboratories, Murray Hill, as the Director of the Dialogue Systems Research Department. Dr. Lee is a Fellow of the IEEE and a Fellow of ISCA. He has published 30 patents and about 600 papers, with more than 55,000 citations and an h-index of about 80 on Google Scholar. He received numerous awards, including the Bell Labs President's Gold Award for speech recognition products in 1998. He won the SPS's 2006 Technical Achievement Award for "Exceptional Contributions to the Field of Automatic Speech Recognition". In 2012 he gave an ICASSP plenary talk on the future of automatic speech recognition. In the same year he was awarded the ISCA Medal in Scientific Achievement for "pioneering and seminal contributions to the principles and practice of automatic speech and speaker recognition". His two pioneering papers on deep regression accumulated over 2000 citations in recent years and won a Best Paper Award from IEEE Signal Processing Society in 2019.



Professor / Editor-In-Chief

Mario Koeppen

Kyushu Institute of Technology, Japan /, Applied Soft Computing

Metaverse as AI Embodiment: Techniques, Impact, and Research Opportunities

Born from the 1992 novel "Snow Crash" of Neal Stephenson as dystopian vision of a society where the depth of technological integration passed a threshold where the human can't distinguish anymore between its real-world identity and digital identity, that thing called metaverse has come into our technical reach. This is due to the rapid advances in sensing, computing, storing, communication and intelligent technologies that more and more thoroughly shape, influence and even condition our all daily life. Metaverse is a global concept and has been characterized by factors such as: identity (avatars), social interaction, immersion, low latency, diversity, anytime anywhere service, trading economy and civilization, to follow organizational principles of a society. It is not restricted to computer games. We can understand, and thus design, optimize, analyze, enrich etc. any large-scaled public service in terms of a metaverse model: transportation networks, smart city, the global food chain, video conferencing systems, to name a few. This way, metaverse also becomes a methodology. In this talk, after introducing above points on a more factual base and showing how well-known public infrastructures can in fact be modeled as a metaverse, and to what that might be good for, we will consider the way of integrating AI and Metaverse. There are essentially three tiers. Tier 1 is the use of AI to design the metaverse. For the artists and developer, the design and functional configuration of digital models targeting a higher and higher level of realism or believability is still a burden, time consuming, experience and skill-based. It is still a highly specific activity where the task can be often only understood in the technical jargon of computer graphics designer. Here, AI already helped to design the tools to create digital assets, and we will show a few examples: auto-posing balanced characters trained by DL, motion extraction from videos, optimization of mesh data controlling animations and scene visualizations, or the creation of compensating dynamic corrections (aka morphs). It is still a highly specific application domain, and for a good reason, in the roadmap of text-based generative AI techniques as there are: text to text, text to image, text to video, text to 3D model, it has been put at the last position. Hope is that we can clarify why that is so. Tier 2, it is the opposite way, the Metaverse assists the development, learning and training of AI and related scientific investigations. This refers to a common understanding of metaverse as simulation and visualization tool. But it's a bit more. It refers to the convergence of parallelization multiple metaverse instances and a continuously increasing level of realism, subjective immersion feeling, and physical correctness with feasible compensations. Most typical case here is the design of training situations for DL training, that often needs a huge corpus of data that can't be created from real-

world situations. Uses are: for object detection from realistically rendered scenes that can be interactively explored, or for challenges to develop robot arm control in a factory environment, or to optimize the control or response to influencing factors in smart farming. Tier 3 is to equip the metaverse with AI. This is a common need in computer games to design the behavior of so-called NPCs (Non-Playable Characters) but otherwise still a vision, and only naive approaches like NPC chatbots have been conceptualized so far. Still, AI is non-divisible, has no identity, and also the idea to give them a body, allow them to communicate, to differentiate, to, well, evolve, to link to real-world processes and take controls might give the one or other the idea that we also should better think twice about what we are doing before we are doing so. But in summary, we are still in a situation where most of the people have no practical experience of a metaverse, and if so hardly find the time needed to spend in it while facing a steep learning curve, apart from simply just not being interested, worried about health impact, diminishing social life activities, become manipulated and distracted from solving the true problems of humankind, it has been shown that the whole concept of metaverse is actually counterfeiting such concerns, and poses a lot of scientific and engineering challenges and opportunities for many years to come.

BIOGRAPHY

Mario Köppen was born in 1964. He studied physics at the Humboldt-University of Berlin and received his master degree in solid state physics in 1991. Afterwards, he worked as scientific assistant at the Central Institute for Cybernetics and Information Processing in Berlin and changed his main research interests to image processing and neural networks. From 1992 to 2006, he was working with the Fraunhofer Institute for Production Systems and Design Technology. He continued his works on the industrial applications of image processing, pattern recognition, and soft computing, esp. evolutionary computation. During this period, he achieved the doctoral degree at the Technical University Berlin with his thesis works: "Development of an intelligent image processing system by using soft computing" with honors. He has published more than 150 peer-reviewed papers in conference proceedings, journals and books and was active in the organization of various conferences as chair or member of the program committee, incl. the WSC on-line conference series on Soft Computing in Industrial Applications, and the HIS conference series on Hybrid Intelligent Systems. He is founding member of the World Federation of Soft Computing, and also Editor of the Applied Soft Computing journal. In 2006, he became JSPS fellow at the Kyushu Institute of Technology in Japan, and in 2008 Professor at the Network Design and Research Center (NDRC) and 2013 Professor at the Graduate School of Creative Informatics of the Kyushu Institute of Technology, where he is conducting now research in the fields of multi-objective and relational optimization, digital convergence and multimodal content management.



Editor-In-Chief

Chunghui Kuo

Journal of Imaging Science and Technology, Society for Imaging Science and Technology

Deep Learning in Image Processing and Computer Vision

The fundamental question for machine learning is to find a rich class of functionals which can map from a high-dimensional space of raw data to a low-dimensional collection of actionable outcomes under the practical constraint of being computationally trackable. Without claiming of its optimality, the Continuous Piecewise Linear (CPL) functionals proposed by the artificial intelligence research community have demonstrated the ability of meeting all requirements listed previously through deep learning network architectures. The capability of analyzing high dimensional data has allowed deep learning to play an indispensable role in image processing and computer vision applications where the number of raw data is extremely high. In this talk, we will first briefly review the theoretical connection between deep learning architecture and Continuous Piecewise Linear functional. Afterwards, popular deep learning networks for classification, detection, segmentation, and generation will be summarized to give the audience an overall picture of deep learning architectures. At last, an outlook of the AI industry will be presented to show projected future growth opportunities and potential technical challenges.

BIOGRAPHY

Chunghui Kuo is the Editor-in-Chief at the Journal of Imaging Science and Technology and founder of Raycers Technology focusing on autonomous printing and intelligent environment sensing.

He received his PhD in Electrical and Computer Engineering from the University of Minnesota and joined Eastman Kodak Company in 2001. He was a senior scientist, a Distinguished Inventor, and served as an IP coordinator at the Eastman Kodak Company. He holds 52 US patents, where his patented automatic white-blending workflow received 2017 Canadian Printing Award. He had served as an industrial standardization committee member from 2005 to 2008 at the International Organization for Standardization (ISO).

He is a senior member of the IEEE Signal Processing Society and his research interests are in image processing, computer vision, blind signal separation and classification, and artificial intelligence applied in signal/image processing.



Professor

Hung-Yi Lee

National Taiwan University, Taiwan

How Far Are We from a Speech Version of ChatGPT?

Over the past few months, the capabilities of large language models like ChatGPT have drawn widespread discussion and amazement, but they only focus on text processing. How far are we from a speech-based ChatGPT? What are we missing that prevents the birth of a speech-based ChatGPT? In this presentation, I will share a series of the latest research findings that might lead us towards a speech-based ChatGPT and explore the current challenges and potential solutions.

BIOGRAPHY

Hung-Yi Lee is an associate professor of the Department of Electrical Engineering of National Taiwan University (NTU), with a joint appointment at the Department of Computer Science & Information Engineering of the university. His recent research focuses on developing technology that can reduce the requirement of annotated data for speech processing (including voice conversion and speech recognition) and natural language processing (including abstractive summarization and question answering). He won Salesforce Research Deep Learning Grant in 2019, AWS ML Research Award in 2020, Outstanding Young Engineer Award from The Chinese Institute of Electrical Engineering in 2018, Young Scholar Innovation Award from Foundation for the Advancement of Outstanding Scholarship in 2019, Ta-You Wu Memorial Award from Ministry of Science and Technology of Taiwan in 2019, and The 59th Ten Outstanding Young Person Award in Science and Technology Research & Development of Taiwan. He owns a YouTube channel teaching deep learning in Mandarin with about 100k Subscribers.



Application Engineer

Neal Huang

Analog Devices, Inc., Taiwan

Energy-Efficient Generative AI Future: Low-Power Applications of ADI Max78000

The ADI Max78000 is a powerful, low-power microcontroller with advanced features like the Arm Cortex-M4F processor and a robust neural network accelerator, enabling efficient execution of complex AI tasks. On the other hand, Generative AI, particularly Generative Adversarial Networks (GANs), can locally generate high-quality images, sounds, and text, adding rich interactivity and creativity to edge devices.

With its low-power characteristics, the ADI Max78000 is perfectly suited for edge devices where energy efficiency is crucial for prolonged operation without frequent recharging or replacing batteries. By carefully optimizing power consumption during AI model inference and leveraging the CNN accelerator, the Max78000 can achieve remarkable AI performance while conserving energy, making it an excellent choice for battery-powered and energy-sensitive applications.

In conclusion, by harnessing the strengths of CNN acceleration and low-power characteristics, the integration of ADI Max78000 with Generative AI unlocks new possibilities for innovative and energy-efficient edge AI applications. Whether it's enhancing image synthesis, generating realistic content, or enabling interactive voice assistants, this powerful combination opens up endless opportunities to create intelligent and sustainable AI solutions for the future.

BIOGRAPHY

Currently, as a part of Analog Devices Inc. (ADI), a world-renowned semiconductor leader, I am thrilled to be a member of the dynamic team responsible for AI-related product research and development. Drawing upon my over 20 years of experience in semiconductor industry, I will bring a wealth of expertise in engineering and FAE roles.

ADI is renowned for its world-class analog, mixed-signal, and digital signal processing (DSP) technologies, which enable the development of a wide range of applications across various sectors. From healthcare and automotive to industrial, communications, and consumer electronics, ADI's products play a crucial role in shaping the future of these industries.

At ADI, we are committed to pushing the boundaries of AI technology and its applications across various industries. Leveraging my extensive background in programming software, firmware, and hardware circuit design, I am excited to contribute to the creation of cutting-edge AI solutions that cater to our customers' specific needs and drive technological advancements in the semiconductor field.