IEEE Signal Processing Society, Gujarat Chapter

Jointly with

Institute of Technology, Nirma University, Ahmedabad

cordially invites you to the expert lecture under

Distinguished Lecture Programme on

Audio-Visual Voice Activity Detection Using Deep Neural Networks

By Dr. Israil Cohen, Professor, Electrical engineering at the Technion – Israel Institute of Technology, Haifa, Israel

Date: Wednesday, March 11, 2020

Venue: A-101, A-Block, Nirma University Campus

Time: 11:00 am to 01:00 pm

Abstract of the Talk:

Recently, there has been growing use of deep neural networks in many modern speech-based systems such as speaker recognition, speech enhancement, and emotion recognition. In this talk, we address the problem of audio-visual voice activity detection (VAD) by incorporating auditory and visual modalities into an end-to-end deep neural network. We evaluate the proposed system in challenging acoustic environments including high levels of noise and transients, which are common in real-life scenarios. Our multimodal setting includes a speech signal captured by a microphone and a corresponding video signal capturing the speaker's mouth region. Under such difficult conditions, robust features need to be extracted from both modalities to accurately distinguish between speech and noise. For this purpose, we utilize a deep residual network, to extract features from the video signal, while for the audio modality; we employ a variant of WaveNet encoder for feature extraction. The features from both modalities are fused using multimodal compact bilinear pooling to form a joint representation of the speech signal. To further encode the temporal information, we feed the fused signal to a long short-term memory network and the system is then trained in an end-to-end supervised fashion. Experimental results demonstrate the improved performance of the proposed end-to-end multimodal architecture compared to unimodal variants for VAD.

Bio Sketch of Dr. Israil Cohen:

Israel Cohen is a Professor of electrical engineering at the Technion - Israel Institute of Technology, Haifa, Israel. He is also a Visiting Professor at Northwestern Polytechnical University, Xi'an, China. He is an IEEE Fellow "for contributions to the theory and application of speech enhancement", and Distinguished Lecturer of the IEEE Signal Processing Society. He received the B.Sc. (Summa Cum Laude), M.Sc. and Ph.D. degrees in electrical engineering from the Technion - Israel Institute of Technology, in 1990, 1993 and 1998, respectively.

From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT, USA. In 2001 he joined the Electrical Engineering Department of the Technion.

He is a coeditor of the Multichannel Speech Processing Section of the Springer Handbook of Speech Processing (Springer, 2008), and a coauthor of Fundamentals of Signal Enhancement and Array Signal Processing (Wiley-IEEE Press, 2018).

He was awarded the Norman Seiden Prize for Academic Excellence (2017), the SPS Signal Processing Letters Best Paper Award (2014), the Alexander Goldberg Prize for Excellence in Research (2010), and the Muriel and David Jacknow Award for Excellence in Teaching (2009).

He served as Associate Editor, IEEE Transactions on Audio, Speech, and Language Processing (2004-2007); Associate Editor, IEEE Signal Processing Letters (2004-2008); Member, Audio and Acoustic Signal Processing Technical Committee (2012-2017); and Member, Speech and Language Processing Technical Committee (2013-2015).

His research interests are array processing, statistical signal processing, deep learning, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification and adaptive filtering.

For More detail resume of the speaker, you can visit his home page link- https://israelcohen.com/

E-mail: icohen@ee.technion.ac.il

For any Details Contact:

Dr. Tanish Zaveri, Professor –EC Dept., SOT, IT-NU

Ph. No. 9824803401.