Title: Deep Scattering End-to-End Architectures for Speech Recognition

Authors: Iyalla John Alamina (1), David Wilson (1), Andrew Crampton (1)

Affiliation: (1) University of Huddersfield

Email: john.alamina@hud.ac.uk, [d.r.wilson@hud.ac.uk](mailto:d.r.wilson@hud.ac.uk),

Format: Talk

Abstract

This work explores the prospects of deep recurrent end-to-end architectures applied to speech recognition. Complementary aspects of developing speech recognition systems are eliminated by focusing on end-to-end speech units as a two-step process requiring a Connectionist Temporal Character Classification (CTCC) model and Language Model (LM) rather than a three-step process requiring an Acoustic model(AM), LM and phonetic dictionary. A two-step process rather than a three-step process is particularly desirable for low resource languages as resources are required to build only two models instead of three models.

Our Bi-directional Recurrent neural network (Bi-RNN) end-to-end system, is augmented by features derived from a deep scattering network as opposed to the standard Mel Cepstral (MFCC) features used in state of the art acoustic models. These specialised deep scattering features, consumed by the Bi-RNN, model a light-weight convolution network. This work shows that it is possible to build a speech model from a combination of deep scattering features and a Bi-RNN. There has been no record of deep scattering features being used in end-to-end bi-RNN speech models as far as we are aware.