BAI Proposal - Unifying Supervised and Unsupervised Deep Representation Learning of Audio-Visual Signals

Sriram Ganapathy (EE) & Prathosh A. P. (ECE)

Summary

The application of deep learning to generate meaningful representations from data, termed as deep representation learning [1], has received widespread interest in the last decade owing to the advancements in deep learning methods. This area initially started with successful approaches for supervised representation learning in language and vision tasks. In recent years, unsupervised representation learning has also been explored with objectives of self-supervision [2] and clustering. In this project, the goal is to develop representations that are jointly optimal for multiple data modalities with or without supervision. One of the key objectives of representation learning is the ability to disentangle the factors that describe the data. This will also allow data representations to succinctly reconstruct the data and generate a subset of the factors than can categorize the data to classes. The ideas will also be investigated on 1-D signals like speech/audio as well as 2-D visual signals.

The success of the representations will be judged both for supervised tasks as well as for data generation tasks. The data generation tasks involve sampling new data points from unknown distributions to generate realistic data while supervised tasks involve classification settings. The project will also make systematic comparisons between representations learned by deep networks with representations observed in brain recordings (using publicly available data).

Example applications include audio zooming where audio source separation is carried out based on visual cues from a video, behavioural assessment of children with Autism using interactive multimodal audio-visual data, summary extraction from sports videos, missing video-frame prediction, video generation etc.

Student Qualifications

The successful student who would work on this problem would be the one with an engineering background in the streams of Electrical, Electronics or Computer Engineering as well as those with mathematical/biological sciences background with an inclination to understand human and machine representations.

References

[1] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8,pp. 1798–1828, 2013.

[2] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). “A simple framework for contrastive learning of visual representations.” In International conference on machine learning (pp. 1597-1607). PMLR.