

8.

My recent research interest is sound source localization for unknown number of sources. To start with, sound source localization (SSL) is the technology of estimating the Direction-of-Arrival (DOA) of one or several sound sources from the multichannel signals captured by the microphone array. Most methods have been focusing on localizing a single sound source, which do not extend to multiple sources rather than unknown number of sources. However, to estimate the DOA of unknown number of sources is the real key challenge, cause the number of sources is often unknown in real applications and SSL's performance decrease sharply in unknown number of speakers condition.

The first paper I read is *DNN for Multiple Speaker Detection and Localization*. They propose a likelihood-based encoding of the network output so that the spatial pseudo-spectra (SPS) can be generated by the network and then select form peaks of the SPS that are above a certain threshold when the number of sound sources is unknown. The idea of generating SPS by the network can work and it is no worse than that generated by conventional methods. It does achieve relatively good results of localization with a unknown number of sources, but the threshold is hard to choose and a pre-determined threshold for all domains are prone to errors due to the domain shift problem.

The second one is *Robust Source Counting and DOA Estimation Using SPS and CNN*. They propose to use a CNN to estimate the number of sources from short-time SPS. And the estimated number of sources is used to select the final DOAs without using a fixed threshold. The idea of firstly estimating number of sources is fancy, but the accuracy of the estimation is not high enough (around 70%), thus misleading to wrong DOA estimation. In the future, I'd like to do more research on improving the accuracy of estimating the number of sources, proposing a more efficient network combined with the SPS estimating network mentioned above, thus can do more accurate DOA estimation for an unknown number of sources in an end-to-end fashion.