[7]

In Adavanne et al, LSTM units (Long short-term memory) within a Recurrent Neural Network (RNN) are leveraged to identify sound events in polyphonic audio samples [7]. Three different feature sets were studied: log mel-band energy, harmonic features (such as pitch), and time difference of arrival (TDOA). A combination of mel-band energy and TDOA features performed the best in this context.

[8] and [9]

Royo-Letelier et al [8] explored the use of metric learning to disambiguate artists within a music catalog. They found that at smaller dataset sizes (300 or fewer examples), that a traditional 1D-CNN (One Dimensional Convolutional Neural Network) outperformed the metric learning model [8]. The model they compared their metric learning model to is described in [9].

In [8], triplet loss was used to separate different groups of musicians. This loss mechanism learns a metric preserving map *f* that groups similar pairs and distances dissimilar pairs given a triplet of (xa, x+, x-).

# References

|  |  |
| --- | --- |
| [1] | Federal Aviation Administration, "Aircraft Noise Issues," [Online]. Available: https://www.faa.gov/about/office\_org/headquarters\_offices/apl/noise\_emissions/airport\_aircraft\_noise\_issues/. [Accessed 1 12 2018]. |
| [2] | US National Parks Service, "Sounds Gallery," [Online]. Available: https://www.nps.gov/subjects/sound/gallery.htm. [Accessed 2 12 2018]. |
| [3] | Y. Wang, X. Wu, X. Li and J. Zhou, "Vehicle type recognition in sensor networks using improved time encoding signal processing algorithm," *Mathematical Problems in Engineering,* vol. 2014, 2014. |
| [4] | H. Wu, M. Siegel and P. Khosla, "Vehicle sound signature recognition by frequency vector principal component analysis," in *Instrumentation and Measurement Technology Conference*, 1998. |
| [5] | J. Andén and S. Mallat, "Deep scattering spectrum," *IEEE Transactions on SIgnal Processing,* vol. 62, no. 16, pp. 4114-4128, 2014. |
| [6] | F. Hemmati, W. Orfali and M. Gadala, "Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation," *Applied Acoustics,* vol. 104, pp. 101-118, 2016. |
| [7] | S. Adavanne, G. Parascandolo, P. Pertilä, T. Heittola and T. Virtanen, "Sound event detection in multichannel audio using spatial and harmonic features," in *Detection and Classification of Acoustic Scenes and Events*, Budapest, Hungary, 2016. |
| [8] | J. Royo-Letelier, R. Hennequin, V. Tran and M. Moussallam, "Disambiguating music artists at scale with audio metric learning," in *19th International Society for Music Information Retrieval Conference*, Paris, 2018. |
| [9] | J. Park, J. Lee, J. Park, J. Ha and J. Nam, "Representation learning of music using artist labels," in *19th International Society for Music Information Retrieval Conference*, Pairs, 2018. |
| [10] | Atlassian, "Agile Tools," [Online]. Available: https://trello.com/power-ups/59d4ef8cfea15a55b0086614/agile-tools. [Accessed 1 12 2018]. |