Resources for Audio Event Detection & Scene Analysis w/ ML

by Fabian Seipel, Corvin Jaedicke

September 2, 2020

1. Audio DSP

Tutorials/Courses:

- https://www.coursera.org/learn/audio-signal-processing
- https://www.audiolabs-erlangen.de/resources/MIR/FMP/C0/C0.html
- https://www.youtube.com/watch?v=rkvEM5Y3N60&list=PL8157CA8884571BA2
  [MIT RES.6-008 Digital Signal Processing, 1975]
- https://github.com/mgeier/signals-and-systems-lecture
- https://github.com/mgeier/python-audio
- https://www.ak.tu-berlin.de/fileadmin/a0135/Unterrichtsmaterial/Weinzierl/EDS%20Skript.pdf

Literature:

- Schafer, R.W. and Oppenheim, A.V., 1989. Discrete-time signal processing. Englewood Cliffs, NJ: Prentice Hall.
- Weinzierl, S. ed., 2008. Handbuch der Audiotechnik. Springer Science & Business Media.
- Muller, M., Ellis, D.P., Klapuri, A. and Richard, G., 2011. Signal processing for music analysis. IEEE Journal of Selected Topics in Signal Processing, 5(6), pp.1088-1110.
- Lerch, A., 2012. An introduction to audio content analysis: Applications in signal processing and music informatics. Wiley-IEEE Press.

2. Deep Learning

Tutorials/Courses:

- https://colab.research.google.com/drive/1UCJt8EYjlzCs1H1d1X0iDGYJsHKwu-NO
- https://developers.google.com/machine-learning/crash-course
- https://www.coursera.org/learn/machine-learning
- https://www.coursera.org/specializations/deep-learning
- https://www.udacity.com/course/intro-to-tensorflow-for-deep-learning–ud187
- https://www.udacity.com/course/intro-to-tensorflow-lite–ud190
- http://cs229.stanford.edu/
- http://cs231n.stanford.edu/

Literature:

- LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), p.436.
- Goodfellow, I., Bengio, Y. and Courville, A., 2016. Deep learning. MIT press.

- Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.

3. Acoustic Event Detection & Scene Classification

Conferences:

- http://dcase.community/workshop2019/
- https://2020.ieeeicassp.org/

Datasets:

- http://www.cs.tut.fi/~heittolt/datasets [Collection]
- https://github.com/ybayle/awesome-deep-learning-music/blob/master/datasets.md
- https://annotator.freesound.org/fsd/downloads/ [FSD]
- https://zenodo.org/record/3384388#.XaXsWeczbOQ [MIMII Dataset]

Most often used in research:

- http://dcase.community/challenge2018/task-acoustic-scene-classification#audio-dataset
- https://urbansounddataset.weebly.com/urbansound8k.html
- https://github.com/karoldvl/ESC-50
- https://github.com/karoldvl/ESC-10

Python Packages:

- Tensorflow/Keras
- Sklearn
- Pandas
- Librosa
- Madmom

Lectures/Presentations:

- https://www.youtube.com/watch?v=7B1WBa3sC3I
  [Deep Neural Networks for Sound Event Detection]
- https://www.youtube.com/watch?v=zvccOFz2KxI
  [Robust Sound Event Detection in Acoustic Sensor Networks]
- https://www.youtube.com/watch?v=9X66iwEQSyI
  [Audio Event Detection w/Deep Learning at at Stanley Black & Decker]
- https://www.youtube.com/watch?v=IzRCC2-IBwU
  [Machine Listening of Everyday Soundscapes]

Publications/Ressources:

Overview of Acoustic Event Detection [AED] and Scene Classification [ASC]:

- Virtanen, T., Plumbley, M.D. and Ellis, D. eds., 2018. Computational analysis of sound scenes and events. Heidelberg: Springer.

More compact Literature Review Deep Learning for Audio:

- Choi, K., Fazekas, G., Cho, K. and Sandler, M., 2017. A tutorial on deep learning for music information retrieval.
- Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S., Sainath, T.N., 2019. Deep Learning for Audio Signal Processing. IEEE Journal of Selected Topics in Signal Processing 13, 206–219.

Relevant PhD Theses:

- Dieleman, S., 2015. Learning feature hierarchies for musical audio signals (Doctoral dissertation, Ghent University).
- Schlüter, J., 2017. Deep learning for event detection, sequence labelling and similarity estimation in music signals. Ph. D. thesis.
- Kumar, A., 2018. Acoustic Intelligence in Machines, Doctoral dissertation. School of Computer Science: Carnegie Mellon University.
- Cakir, E., 2019. Deep Neural Networks for Sound Event Detection.

Data Augmentation in time-domain:

- Salamon, J. and Bello, J.P., 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Processing Letters, 24(3), pp.279-283.

Data Augmentation in frequency-domain:

- Park, D.S., Chan, W., Zhang, Y., Chiu, C.C., Zoph, B., Cubuk, E.D. and Le, Q.V., 2019. Specaugment: A simple data augmentation method for automatic speech recognition.
- Zhang, Z., Xu, S., Cao, S. and Zhang, S., 2018, November. Deep convolutional neural network with mixup for environmental sound classification. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV) (pp. 356-367). Springer, Cham.

Data Augmentation with Adverserial Networks:

- Donahue, C., McAuley, J. and Puckette, M., 2018. Adversarial audio synthesis.
  [https://github.com/chrisdonahue/wavegan]
  [https://chrisdonahue.com/wavegan_examples/]
  [https://www.youtube.com/watch?v=BA-Z0KJIyJs]
- Engel, J., Agrawal, K.K., Chen, S., Gulrajani, I., Donahue, C. and Roberts, A., 2019. Gansynth: Adversarial neural audio synthesis.
- Unknown Authors, DDSP: Differentiable Digital Signal Processing
  [https://openreview.net/forum?id=B1x1ma4tDr]
- Yamamoto, R., Song, E. and Kim, J.M., 2019. Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram. arXiv preprint arXiv:1910.11480. [https://r9y9.github.io/demos/projects/icassp2020/]

End-to-end learning on audio:

- Dieleman, S. and Schrauwen, B., 2014, May. End-to-end learning for music audio. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6964-6968). IEEE.
- Pons, J., Nieto, O., Prockup, M., Schmidt, E., Ehmann, A. and Serra, X., 2017. End-to-end learning for music audio tagging at scale.
- Abdoli, S., Cardinal, P. and Koerich, A.L., 2019. End-to-End Environmental Sound Classification using a 1D Convolutional Neural Network. Expert Systems with Applications.

Relevant milestone paper:

- Cakır, E., Parascandolo, G., Heittola, T., Huttunen, H. and Virtanen, T., 2017. Convolutional recurrent neural networks for polyphonic sound event detection. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(6), pp.1291-1303.
- Kumar, A., Khadkevich, M. and Fügen, C., 2018, April. Knowledge transfer from weakly labeled audio using convolutional neural network for sound events and scenes. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 326-330). IEEE.
- McFee, B., Salamon, J. and Bello, J.P., 2018. Adaptive pooling operators for weakly labeled sound event detection. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 26(11), pp.2180-2193.
- Pons, J., Serrà, J. and Serra, X., 2019, May. Training neural audio classifiers with few data. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 16-20). IEEE.
- Cramer, J., Wu, H.H., Salamon, J. and Bello, J.P., 2019, May. Look, Listen, and Learn More: Design Choices for Deep Audio Embeddings. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3852-3856). IEEE.

Most recent approaches for AED/ASC:

- http://dcase.community/challenge2019/