References

[1] Sharath Adavanne, Giambattista Parascandolo, Pasi Pertilä, Toni Heittola, and Tuomas Virtanen. Sound event detection in multichannel audio using spatial and harmonic features. *arXiv preprint arXiv:1706.02293*, 2017.

[2] Emmanouil Benetos, Grégoire Lafay, Mathieu Lagrange, and Mark D Plumbley. Detection of overlapping acoustic events using a temporally-constrained probabilistic model. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6450–6454. IEEE, 2016.

[3] Emre Cakır, Giambattista Parascandolo, Toni Heittola, Heikki Huttunen, and Tuomas Virtanen. Convolutional recurrent neural networks for polyphonic sound event detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(6):1291–1303, 2017.

[4] Woohyun Choi, Jinsang Rho, David K Han, and Hanseok Ko. Selective background adaptation based abnormal acoustic event recognition for audio surveillance. In *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, pages 118–123. IEEE, 2012.

[5] Qi Li, Huadong Ma, and Dong Zhao. A neural network based framework for audio scene analysis in audio sensor networks. In *Pacific-Rim Conference on Multimedia*, pages 480–490. Springer, 2009.

[6] Theodoros Giannakopoulos, Alexandros Makris, Dimitrios Kosmopoulos, Stavros Perantonis, and Sergios Theodoridis. Audio-visual fusion for detecting violent scenes in videos. In *Hellenic conference on artificial intelligence*, pages 91–100. Springer, 2010.

[7] Asma Rabaoui, Zied Lachiri, and Noureddine Ellouze. Using hmm-based classifier adapted to background noises with improved sounds features for audio surveillance application. *Int. J. Signal Process*, 3:535–545, 2009.

[8] Stavros Ntalampiras, Ilyas Potamitis, and Nikos Fakotakis. On acoustic surveillance of hazardous situations. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 165–168. IEEE, 2009.

[9] Burak Uzkent, Buket D Barkana, and Hakan Cevikalp. Non-speech environmental sound classification using svms with a new set of features. *International Journal of Innovative Computing, Information and Control*, 8(5):3511–3524, 2012.

[10] Xiaodan Zhuang, Xi Zhou, Mark A Hasegawa-Johnson, and Thomas S Huang. Real-world acoustic event detection. *Pattern Recognition Letters*, 31(12):1543–1551, 2010.

[11] Sharath Adavanne, Pasi Pertilä, and Tuomas Virtanen. Sound event detection using spatial features and convolutional recurrent neural network. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 771–775. IEEE, 2017.

[12] Victor Bisot, Romain Serizel, Slim Essid, and Gaël Richard. Acoustic scene classification with matrix factorization for unsupervised feature learning. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 6445–6449. IEEE, 2016.

[13] Vincenzo Carletti, Pasquale Foggia, Gennaro Percannella, Alessia Saggese, Nicola Strisciuglio, and Mario Vento. Audio surveillance using a bag of aural words classifier. In *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 81–86. IEEE, 2013.

[14] Kristen Grauman and Trevor Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1458–1465. IEEE, 2005.

[15] Anurag Kumar, Pranay Dighe, Rita Singh, Sourish Chaudhuri, and Bhiksha Raj. Audio event detection from acoustic unit occurrence patterns. In

2012 IEEE international conference on acoustics, speech and signal processing (ICASSP), pages 489–492. IEEE, 2012.

[16] Donatello Conte, Pasquale Foggia, Gennaro Percannella, Alessia Saggese, and Mario Vento. An ensemble of rejecting classifiers for anomaly detection of audio events. In *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, pages 76–81. IEEE, 2012.

[17] Michele Lai Chin and Juan José Burred. Audio event detection based on layered symbolic sequence representations. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1953–1956. IEEE, 2012.

[18] Sam T Roweis. One microphone source separation. In *Advances in neural information processing systems*, pages 793–799, 2001.

[19] Laurent Benaroya, Frédéric Bimbot, and Rémi Gribonval. Audio source separation with a single sensor. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):191–199, 2005.

[20] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1–13, 2007.

[21] Konstantinos Trohidis, Grigorios Tsoumakas, George Kalliris, and Ioannis P Vlahavas. Multi-label classification of music into emotions. In *ISMIR*, volume 8, pages 325–330, 2008.

[22] Forrest Briggs, Balaji Lakshminarayanan, Lawrence Neal, Xiaoli Z Fern, Raviv Raich, Sarah JK Hadley, Adam S Hadley, and Matthew G Betts. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131(6):4640–4650, 2012.

[23] Luigi Gerosa, Giuseppe Valenzise, Marco Tagliasacchi, Fabio Antonacci, and Augusto Sarti. Scream and gunshot detection in noisy environments. In *2007 15th European Signal Processing Conference*, pages 1216–1220. IEEE, 2007.

[24] Giambattista Parascandolo, Heikki Huttunen, and Tuomas Virtanen. Recurrent neural networks for polyphonic sound event detection in real life recordings. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6440–6444. IEEE, 2016.

[25] Naoya Takahashi, Michael Gygli, Beat Pfister, and Luc Van Gool. Deep convolutional neural networks and data augmentation for acoustic event detection. *arXiv preprint arXiv:1604.07160*, 2016.

[26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[27] Jiajun Wu, Yinan Yu, Chang Huang, and Kai Yu. Deep multiple instance learning for image classification and auto-annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[28] Huy Phan, Marco Maaß, Radoslaw Mazur, and Alfred Mertins. Random regression forests for acoustic event detection and classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(1):20–31, 2014.

[29] Justin Salamon, Christopher Jacoby, and Juan Pablo Bello. A dataset and taxonomy for urban sound research. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1041–1044, 2014.

[30] Justin Salamon, Duncan MacConnell, Mark Cartwright, Peter Li, and Juan Pablo Bello. Scaper: A library for soundscape synthesis and augmentation. In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 344–348. IEEE, 2017.

[31] Peter Foster, Siddharth Sigtia, Sacha Krstulovic, Jon Barker, and Mark D Plumbley. Chime-home: A dataset for sound source recognition in a domestic

environment. In *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–5. IEEE, 2015.

[32] Emre Cakir, Toni Heittola, Heikki Huttunen, and Tuomas Virtanen. Polyphonic sound event detection using multi label deep neural networks. In *2015 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE, 2015.

[33] M. Cartwright P. Li J. Salamon, D. MacConnell and J. P. Bello. The URBAN-SED Dataset, year = 2017, url = http://urbansed.weebly.com.

[34] Elmar Messner, Matthias Zöhrer, and Franz Pernkopf. Heart sound segmentation—an event detection approach using deep recurrent neural networks. *IEEE transactions on biomedical engineering*, 65(9):1964–1974, 2018.

[35] Justin Salamon and Juan Pablo Bello. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3):279–283, 2017.

[36] Jan Schlüter and Thomas Grill. Exploring data augmentation for improved singing voice detection with neural networks. In *ISMIR*, pages 121–126, 2015.

[37] Wenbo Wang, Sichun Li, Jianshe Yang, Zhao Liu, and Weicun Zhou. Feature extraction of underwater target in auditory sensation area based on mfcc. In *2016 IEEE/OES China Ocean Acoustics (COA)*, pages 1–6. IEEE, 2016.

[38] Stanley A Gelfand. *Hearing: An introduction to psychological and physiological acoustics*. CRC Press, 2016.

[39] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, 2015.

[40] Trevor R Agus, Simon J Thorpe, Clara Suied, and Daniel Pressnitzer. Characteristics of human voice processing. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pages 509–512. IEEE, 2010.

[41] Petros S Karvelis, Dimitrios I Fotiadis, Dimitrios G Tsalikakis, and Ioannis A Georgiou. Enhancement of multichannel chromosome classification using a region-based classifier and vector median filtering. *IEEE Transactions on Information Technology in Biomedicine*, 13(4):561–570, 2009.

[42] S Esakkirajan, T Veerakumar, Adabala N Subramanyam, and CH PremChand. Removal of high density salt and pepper noise through modified decision based unsymmetric trimmed median filter. *IEEE Signal processing letters*, 18(5):287–290, 2011.