# 参考文献

Eronen A J, Peltonen V T, Tuomi J T, et al. Audio-based context recognition[J]. IEEE Transactions on Audio Speech & Language Processing, 2006, 14(1):321-329.

Malkin R G, Waibel A. Classifying user environment for mobile applications using linear autoencoding of ambient audio[C]// ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing. 2005.

Krijnders J, t Holt G. Tone-fit and MFCC scene classification compared to human recognition[J]. Energy [dB], 2013, 400(450): 500.

Sawhney N, Maes P. Situational awareness from environmental sounds[J]. Tech-nical Report, Massachusetts Institute of Technology, 1997.

Clarkson B, Sawhney N, Pentland A. Auditory context awareness via wearable computing[J]. Energy, 1998, 400(600): 20.

Patil K, Elhilali M. Multiresolution auditory representations for scene classification[J]. cortex, 2002, 87(1): 516-527.

Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

Piczak K J. Environmental sound classification with convolutional neural networks[C]//2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2015: 1-6.

Ballas J A. 1993. Common factors in the identification of an assortment of brief everyday sounds[J]. Journal of experimental psychology: human perception and performance, 19(2): 250.

Peltonen V T K, Eronen A J, Parviainen M P, et al. 2001. Recognition of everyday auditory scenes: potentials, latencies and cues[J]. PREPRINTS-AUDIO ENGINEERING SOCIETY.

Dubois D, Guastavino C, Raimbault M. 2006. A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories[J]. Acta acustica united with acustica, 92(6): 865-874.

Tardieu J, Susini P, Poisson F, et al. 2008. Perceptual study of soundscapes in train stations[J]. Applied Acoustics, 69(12): 1224-1239.

Eronen A, Tuomi J, Klapuri A, et al. 2003. Audio-based context awareness-acoustic modeling and perceptual evaluation[C]//Acoustics, Speech, and Signal Processing, 2003.

Lin M, Chen Q, Yan S. 2013. Network in network[J]. arXiv preprint arXiv:1312.4400.

Davis S，Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences [J]. IEEe transactions on acoustics, speech，and signal processing, 1980, 28 (4): 357-366.

Lidy T, Schindler A. 2016. CQT-based convolutional neural networks for audio scene classification[C]//Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 Workshop (DCASE2016). DCASE2016 Challenge, 90: 1032-1048.

Aytar Y, Vondrick C, Torralba A. 2016. Soundnet: Learning sound representations from unlabeled video[C]//Advances in Neural Information Processing Systems. 892-900.

Santoso A, Wang C Y, Wang J C. Acoustic scene classification using network-in-network based convolutional neural network[R]. DCASE2016 Challenge, Tech. Rep, 2016.

Reynolds D A, Rose R C. Robust text-independent speaker identification using Gaussian mixture speaker models[J]. IEEE transactions on speech and audio processing, 1995, 3(1): 72-83.

Dempster A . Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society, Series B, 1977, 39.

Nuttall A H . Some Integrals Involving the Q-Function[J]. IEEE Transactions on Information Theory, 1972, 21(1):95-96.

Rijsbergen C J V. Information Retrieval[M]. 1979.

Wiesel T N , Hubel D H . EXTENT OF RECOVERY FROM THE EFFECTS OF VISUAL DEPRIVATION IN KITTENS[J]. Journal of Neurophysiology, 1965, 28(6):1060-1072.

Fukushima K. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position[J]. Biological Cybernetics, 1980, 36(4): 193-202.

Hinton G E, Osindero S, Teh Y W, et al. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

Salakhutdinov R, Mnih A, Hinton G E, et al. Restricted Boltzmann machines for collaborative filtering[C]. international conference on machine learning, 2007: 791-798.

Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.

Shimodaira H. Improving predictive inference under covariate shift by weighting the log-likelihood function[J]. Journal of statistical planning and inference, 2000, 90(2): 227-244.

|  |
| --- |
|  |

Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv preprint arXiv:1207.0580, 2012.

周志华. 机器学习[M]. 清华大学出版社, 2016.

Eghbal-Zadeh H, Lehner B, Dorfer M, et al. CP-JKU submissions for DCASE-2016: a hybrid approach using binaural i-vectors and deep convolutional neural networks[J]. IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE), 2016.

Ono N, Miyamoto K, Le Roux J, et al. Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram[C]//2008 16th European Signal Processing Conference. IEEE, 2008: 1-4.

Park J, Shin J, Lee K. Exploiting continuity/discontinuity of basis vectors in spectrogram decomposition for harmonic-percussive sound separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(5): 1061-1074.

Moore A W, Jorgenson J W. Median filtering for removal of low-frequency background drift[J]. Analytical chemistry, 1993, 65(2): 188-191.

Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

Wolpert D H. Stacked generalization[J]. Neural networks, 1992, 5(2): 241-259.

Van der Laan M J, Polley E C, Hubbard A E. Super learner[J]. Statistical applications in genetics and molecular biology, 2007, 6(1).

# 致 谢