

- IEEE声学、语音和信号国际会议《加工》(ICASSP), 2019年, 第3135页。
- [20] 孔, 余, 徐, 伊克巴尔, 王, 普拉布利, 弱IEEE/ACM的注意力神经网络标记音频集《音频、语音和语言处理学报》, 第27卷, pp. 17911802, 2019。
- [21] A. Van Den Oord, S. Dieleman和B. Schrauwen, 迁移学习 by supervised pre-training for audio-based music classification,” in 国际音乐信息检索学会会议 (ISMIR), 2014年, 第2934页。
- [22] 王, 弱标记复音事件检测博士论文, 卡内基梅隆大学, 2018年。
- [23] E. Law和L. Von Ahn, 投入协议: 一种新的机制使用人类计算游戏收集数据, 发表在《美国科学院院刊》上 2009年SIGCHI计算系统人为因素会议, 第11971206页。
- [24] A. Mesaros, T. Heittola和T. Virtanen, TUT声学数据库欧洲会议上的场景分类和声音事件检测-pear信号处理会议 (EUSIPCO), 2016, 第11281132页。
- [25] J. Pons和X. Serra, MUSICNN: 预训练卷积神经网络音乐音频标签网络” arXiv预印本arXiv:1909.06654, 2019。
- [26] A. Diment和T. Virtanen, “弱标记音频的迁移学习”在IEEE音频和音频信号处理应用研讨会上声学 (WASPAA), 2017, 第610页。
- [27] 李, 塞西, 迪米特罗娃, 麦基, 《一般分类》用于基于内容的检索的音频数据, “模式识别字母, 第22卷, 第533544页, 2001年。
- [28] L. Vuegen, B. Broeck, P. Karsmakers, J. F. Gemmeke, B. Vanrumste, H. Hamme, 一种用于事件检测的MFCC-GMM方法 IEEE信号处理应用研讨会 音频与声学 (WASPAA), 2013年。
- [29] A. Mesaros, T. Heittola, A. Eronen和T. Virtanen, 声学事件欧洲信号处理协会在现实生活记录中的检测-参考文献 (EUSIPCO), 2010年, 第12671271页。
- [30] B. Uzkent, B. D. Barkana和H. Cevikalp, 非言语环境使用具有一组新特征的SVM进行声音分类, Internationa-国家创新计算、信息与控制杂志, 第8卷, 第35113524页, 2012年。
- [31] 戴, 戴, 屈, 李, 达斯, 非常深卷积 IEEE国际会议上的原始波形神经网络 声学、语音和信号处理 (ICASSP), 2017, 第421页 425。
- [32] 何, 张, 任, 孙, 深度残差学习 IEEE计算机视觉与模式会议上的图像识别《认可》(CVPR), 2016年, 第770778页。
- [33] 孔, 曹, 伊克巴尔, 徐, 王, 普拉布利, 用于音频标记、声音事件检测和空间定位: DCASE 2019基线系统, arXiv预印本 arXiv:1904.034762019。
- [34] A. Krizhevsky, I. Sutskever和G. E. Hinton, ImageNet分类-深度卷积神经网络,” in NeurIPS! 信息处理系统 (NeurIPS), 2012, 第10971105页。
- [35] K. Simonyan和A. Zisserman, 非常深的卷积网络大规模图像识别, “国际学习会议 陈述 (ICLR), 2015年。
- [36] S. Ioffe和C. Szegedy, 批处理归一化: 加速深度通过减少内部协变量转换进行网络训练, “在International! 机器学习会议 (ICML), 2015, 第448456页。
- [37] V. Nair和G. E. Hinton, 修正线性单元改善了限制玻尔兹曼机器, 在国际机器学习会议上 (ICML), 2010年, 第807814页。
- [38] N. 斯里瓦斯塔瓦, G. Hinton, A. Krizhevsky, I. Sutskever和R. Salakhutdinov, Dropout: 一种防止神经网络过拟合的简单方法-fitting, 《机器学习研究杂志》, 第15卷, 第1929页 1958, 2014。
- [39] 林, 陈, 严, 网络中的网络, 国际 2014年离职代表会议 (ICLR)
- [40] A. G. 霍华德, M. ZhuB. 陈, D. Kalenichenko, 王, T. Weyand, M. Andreetto和H. Adam, 《移动网络: 高效卷积》-用于移动视觉应用的神经网络解决方案, arXiv预印本 arXiv:1704.048612017。
- [41] 桑德勒, 霍华德, 朱, 日莫吉诺夫和陈, MobileNetV2: IEEE中的逆残差和线性瓶颈 计算机视觉与模式识别会议 (CVPR), 2018, 第45104520页。
- [42] J. Lee, J. Park, K. L. Kim和J. Nam, 样本级深度卷积 Sound中使用原始波形进行音乐自动标记的神经网络 音乐计算会议, 2017, 第220226页。
- [43] J. Salamon, C. Jacoby和J. P. Bello, 城市数据集和分类学声音研究, 发表在ACM国际会议论文集上 多媒体, 2014, 第10411044页。
- [44] 张, 西塞, 多芬和帕兹, 混音: 超越经验风险最小化”, 在国际学习会议上 陈述 (ICLR), 2018年。
- [45] 朴, 陈, 张-C. Chiu, B. Zoph, E. D. Cubuk 和Q. V. Le, \*SpecAugment: 一种简单的数据增广方法 自动语音识别, ” 2019年 INTERSPEECH。
- [46] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, O. Nieto, “librosa: python中的音频和音乐信号分析” 《Python科学会议论文集》, 2015年第8卷, pp. 1825。
- [47] D. P. Kingma和J. Ba, “Adam: 一种随机优化方法, 2015年国际学习表征会议 (ICLR)。
- [48] L. Ford, H. Tang, F. Grondin, J. Glass, 深残差网络对于大规模声学场景分析, INTERSPEECH, 第25682572页, 2019。
- [49] H. B. Sailor, D. M. Agrawal和H. A. Patil, \*无监督滤波器组基于卷积约束玻尔兹曼机的环境学习-心理声音分类, 载于《国际演讲》, 2017年, 第31073111页。
- [50] K. J. Piczak, ESC: 环境声音分类数据集, 载于 ACM国际多媒体会议, 2015, 第10151018页。
- [51] 陈, 刘, 刘, 张, 严 基于多种分类器的声场景数据增强方案 建模, “DCASE2019挑战赛, 技术代表, 2019。
- [52] 郑和林, DCASE 2018的音频标记系统: 专注于标签噪声数据增强及其高效学习, ” DCASE挑战赛技术代表, 2018年。
- [53] T. Chen和U. Gupta, 基于注意力的卷积神经网络对于具有特征转移学习的音频事件分类, ” [https://cvssp.org/projects/making\\_sense\\_of\\_sounds/site/assets!](https://cvssp.org/projects/making_sense_of_sounds/site/assets!) 挑战 摘要 图片/天翔 深圳, pdf, 2018。
- [54] E. Fonseca, M. Plakal, F. Font, D. P. W. Ellis, X. Favory, J. Pons和 X. Serra, 用音频设备对自由声音进行通用标记-bels: 任务描述、数据集和基线, 在检测研讨会上 声学场景和事件分类 (DCASE), 11月 2018年, 第6973页。
- [55] C. 克罗斯, O. Bones, Y. 曹, L. 哈里斯, P. J. 杰克逊, W. J. 戴维斯, 王, 考克斯和普拉布利, “环境的普遍化”-心理声音分类: 声音数据集和 IEEE声学、语音和声学国际会议上的“挑战” 信号处理 (ICASSP), 2019, 第80828086页。
- [56] C. 刘, L. 冯, G. 刘, H. 王, S. 刘, 自下而上-用于音乐流派分类的cast神经网络 arXiv:1901.089282019。
- [57] G. Tzanetakis和P. Cook, 音频的音乐流派分类 信号, IEEE语音和音频处理汇刊, 第10卷, 第293302页, 2002年。
- [58] 曹, 毛, 彭, 易, 基于谱图的多任务 音频分类, “多媒体工具和应用, 第78卷, pp. 37053722, 2019。
- [59] S. R. Livingstone, K. Peck和F. A. Russo, 《RAVDESS: The Ryerson》 年会“情感言语与歌曲视听数据库” 加拿大 行为与认知科学学会 (CSBCCS), 2012年, 第14591462页。