

- Rev. Lett., 2011, 106(21): 213902.
- [22] Mellin S D, Nordin G P. Limits of scalar diffraction theory and an iterative angular spectrum algorithm for finite aperture diffractive optical element design[J]. Opt. Express, 2001, 8(13): 705–722.
- [23] Gerchberg R W, Saxton W O. A practical algorithm for the determination of phase from image and diffraction plane pictures[J]. Optik, 1972, 35(2): 237–250.
- [24] St-Hilaire P. Phase profiles for holographic stereograms[J]. Opt. Eng., 1995, 34(1): 83–90.
- [25] Gallagher N C, Liu B. Method for computing kinoforms that reduces image reconstruction error[J]. Appl. Opt., 1973, 12(10): 2328–2335.
- [26] Fienup J R. Iterative method applied to image reconstruction and to computer-generated holograms[J]. Opt. Eng., 1980, 19(3): 297–305.
- [27] Wyrowski F, Bryngdahl O. Iterative fourier-transform algorithm applied to computer holography[J]. J. Opt. Soc. Am. A., 1988, 5(7): 1058–1065.
- [28] Wyrowski F, Bryngdahl O. Digital holography as part of diffractive optics[J]. Rep. Prog. Phys., 1991, 54(12): 1481–1571.

◇ 声学新闻和动态 ◇

声学场景和事件的检测及分类

随着计算机听觉场景分析领域的发展,声学场景和事件的检测、分类得到了越来越多的关注。DCASE大赛(The Challenge on Detection and Classification of Acoustic Scenes and Events)自2013年起举办,每届都会设置若干个专门针对声学场景和事件的检测、分类的任务,并发布了相关的音频文件库和基线系统,以期促进该领域的发展。

最近,来自芬兰坦佩雷理工大学、伦敦玛丽王后大学、法国南特中央理工大学以及英国萨里大学的研究学者们在IEEE/ACM Transactions on Audio, Speech and Language Processing杂志上发表的Detection and classification of acoustic scenes and events: outcome of the DCASE 2016 challenge一文,针对2016年DCASE大赛的各项参赛作品的设计和性能进行了详细分析,研究发现深度学习的方法已经逐渐取代基于高斯混合模型和支持向量机的传统分类方法,成为了最热门的分类方法。而特征表示方法这些年来并没有很大的变化,主流做法仍然基于梅尔频率。本赛事的部分资料是公开的,因此对于后续的相关研究很有价值。

DCASE 2016大赛包括声学场景分类,合成音频、现实场景中的声学事件检测,以及音频标注四大项目。其中声学场景分类是一种宽泛的分类方式,目的是把来自不同声源或者位置的声音进行区分。目前广泛使用的特征有梅尔频率倒谱系数MFCC、声学事件直方图或基于时频学习的梯度直方图等。而在建模方面,除了经典的统计学模型如隐马尔科夫模型HMM、高斯混合模型GMM、支持向量机SVM以外,深度学习的方法得到越来越广泛的关注。

声学事件分类与检测则更为细致,目的是把来自相同声源的声音或者是经过相同物理过程产生的声音划为一类进行区分。在早期的研究中,人们把分类和检测混为一谈,关注点一般集中在单个孤立声音上。简单的声学场景中每个音频段只包含一个声学事件,且其中的各类声音没有混叠。目前声学事件分类检测一般是以梅尔刻度的谱特征为基础,

采用经典的HMM、非负矩阵分解NMF,或基于深度学习的方法如卷积神经网络CNN以及长短期记忆LSTM等方法进行研究。

声学场景分类和事件检测的另一大研究热点是室内声学事件的标注。我们可以给一个声音片段标注上一个或者多个标签,而不需要去详细定位事件的具体分类。这种标记方式不涉及声学事件的开始和持续时间,因此也被称为弱标记。该手段被广泛应用于音乐信号的分析处理中,但是在环境音频的标注方面尚未得到深入的研究。目前的研究方法主要有GMM,结合多示例学习的SVM,无监督特征学习以及CNN。相比于环境声信号、语音和音乐信号,室内音频信号的获取更为困难。针对这个问题,DCASE大赛专门录制了特定场景的音频数据库,虽然数据量还不是特别大,但涵盖了日常生活的大部分方面,可用于作为交叉训练和验证模型的样本。通过公开数据库,确定普适性的评价标准和过程,并设定基线工具,有利于实现不同算法的独立评估。

2013年,DCASE首届大赛只设置了声学场景分类和声学时间检测两大方向,2016年扩展为四个方向,加入更为细致的声学事件分类与检测,提出的任务更加贴近日常生活,引起更多的关注和参与度。2017年,大赛又根据该领域的最新发展及关注热点,增加了稀有事件检测和无人驾驶中的大规模弱监督声学事件检测。通过比赛任务的设置、样本库的录制以及对参赛作品的评测,可以更好地把握该方向的发展。DCASE大赛也促进了相关领域的会议增加声音分类的相关主题并提供更多的样本库,极大地推进了该方向的进步。

(余紫莹 编译自: Mesaros A, Heittola T, Benetos E, et al. Detection and classification of acoustic scenes and events: outcome of the DCASE 2016 challenge[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2018, 26(2): 379–393.)