基于卷积神经网络的音频场景分类方法研究

**摘要**

1. **绪论**
   1. 课题背景及研究的目的及意义

声音，一种传递信息的媒介。我们能从声音中提取到不少有用的信息。伴随着技术的发展，借助机器辅助人们从声音中提取出信息，为人工智能领域的一大实践。 更具体地说，通过分析声音使设备能够理解其环境是音频场景识别的主要目标。机器听音系统为人类听觉系统执行类似的处理任务，且通过机器学习，机器人技术和人工智能等领域试该主题的研究更进一步。

？？？引出ASC这个领域

ASC的应用场景包括设计上下文感知服务（Adams, Want，1994），智能可穿戴设备（Xu, Li, Lee 2008），机器人导航系统（Chu, Narayanan, Kuo, Matari，2006）和音频归档管理（Landone, Harrop, Reiss，2008）。此外，智能个人助理（IPA）也是一个受到ASC推动的领域。 IPA是通过分析各种输入数据（包括音频，图像，用户输入或位置，天气和个人时间表等上下文信息）自动进行推荐和执行操作的软件代理。 IPA服务，如Google的Google Now、微软的Cortana、Apple的Siri以及亚马逊的Alexa广泛使用音频输入。这些服务从环境音频中提取上下文信息，可以向用户自动推荐具有价值的信息，是一种极具实用价值的人工智能应用。

此前音频场景分类仍然基于将通用分类器（高斯混合模型，支持向量机，隐马尔可夫模型）应用于手动提取的特征，例如梅尔频率倒谱系数。近年来，得益于计算机速度的提升与深度学习的快速发展，人们逐渐意识到，可以尝试用深度学习的自动特征提取的特性来代替以往低效的手工提取。正如“深度学习”一词所表明的那样，该方法通过使用非线性模块堆叠多个层来进行低层数据的高级表示。有几种深度学习体系结构的变体，卷积神经网络（Convolutional Neural Network, CNN）是深度学习技术中的一种，由于其在学习独特的局部特征方面的优越性能，被广泛用于图像分类、语音识别、自然语言处理。与其他深度学习结构相比，卷积神经网络在图像和[语音识别](https://zh.wikipedia.org/wiki/%E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB)方面能够给出更好的结果。相比较其他深度、前馈神经网络，卷积神经网络需要的参数更少，使之成为一种颇具吸引力的深度学习结构。

2015年,Piczak等人（Piczak，2015）对深度学习中的卷积神经网络是否可有效的应用于音频场景分类这一问题进行了探讨。为此他们依照此前将卷积神经网络成功用于图像分类的经验运用于音频场景分类上。实验结果表明，使用卷积神经网络进行音频场景分类是一个切实可行的办法。卷积神经网络模型胜过基于手动设计特征的常用方法，并达到与其他特征学习方法类似的水平。且卷积神经网络即使在有限的数据集和简单的数据增强下也可以有效应用于环境声音分类任务。更重要的是，可用数据集规模的显著增加很可能大大提高训练模型的性能。得益于卷积神经网络对数据集的利用程度高及高效的类别学习特性，可以看出卷积神经网络对音频场景分类任务有很高的价值。

* 1. 国内外研究现状

1.2.1 音频特征提取

* 1. 研究内容

1. **基于 MFCC 和 KNN 的场景识别基线系统**
2. **基于MFCC 和CNN的场景识别系统**
3. **实验结果分析及对比**

**结论**

**参考文献**

**致谢**