



科研进展

您当前的位置: [首页 \(http://www.ioa.cas.cn/\)](http://www.ioa.cas.cn/) > [科研成果 \(../..\)](#) > [科研进展 \(../\)](#)

中国科学院噪声与振动重点实验室Lab9_DSP411团队在ICASSP 2022 L3DAS22 Challenge中夺冠

发布时间: 2022-03-14 作者: 噪声与振动重点实验室

(<http://www.ioa.cas.cn/>)

url=[http://www.ioa.cas.cn/](#)

中国科学院噪声与振动重点实验室Lab9_DSP411团队参加了音频信号处理领域顶级会议ICASSP 2022举办的深度学习3D音频信号处理挑战赛 (L3DAS22: Machine Learning for 3D Audio Signal Processing), 在L3DAS22任务二——3D声事件定位与检测 (Sound Event Localization and Detection, SELD) 中斩获冠军。

%E5%

logo.

2022 IEEE ICASSP Grand Challenge



L3DAS22: Machine Learning for 3D Audio Signal Processing Signal Processing Grand Challenge at IEEE ICASSP 2022

L3DAS22 Challenge是由快手联合意大利罗马第一大学 (Sapienza University of Rome, Italy) 在ICASSP 2022 Grand Challenge中发起的挑战赛。该挑战赛旨在促进3D信号处理的机器学习研究。本次挑战赛共设置两个子任务:

任务一: 多通道3D语音增强任务, 专注于办公室场景的实时语音增强需求。

任务二: 真实场景下特定声事件定位和检测任务。

两项任务都是在混响办公室条件下使用两个一阶全景声麦克风 (First-order Ambisonics, FOA) 录制。

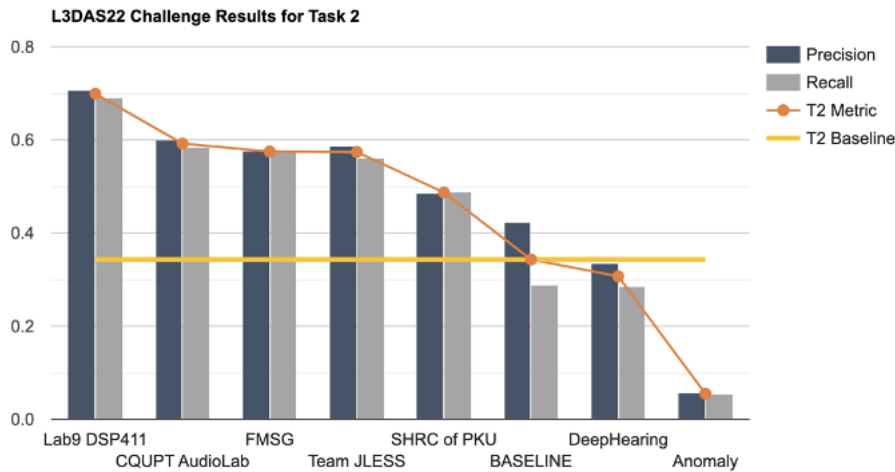
L3DAS22的数据集采用实录的房间脉冲响应 (Room Impulse Responses, RIRs) 与纯净的单声道声源信号卷积合成。在任务二中, 最多有3个声事件同时发生, 而这些声事件可能都属于同一个类别。该任务需要在100毫秒的时间精度下预测这些活动的声事件的类别与其对应的空间位置。该任务使用的是声源定位和声事件检测的联合估计指标: 对定位敏感的检测指标F-score, 定位阈值为2米。在预测正确类别的声事件中, 只有当声定位误差在2m范围内, 该预测才是有效的。

基于深度学习的3D音频信号处理技术引起学术界和工业界广泛关注, 在虚拟和真实会议、游戏开发、音乐制作, 自动驾驶及监控等领域深入应用。相较于单通道音频, 3D音频携带的空间三维声源方位信息, 有助于提升语音和情感识别、声源分离、语音增强和去噪, 以及促进机器声场感知效果。

中国科学院噪声与振动重点实验室的杨军研究员、吴鸣研究员和杨飞然研究员带领博士研究生胡锦涛, 联合英国萨里大学 (University of Surrey) 的Mark D. Plumbley教授、曹寅博士和孔秋强博士 (现任职于字节跳动) 组成Lab9_DSP411团队参加此次比赛, 取得L3DAS22 Task 2中第一名的成绩。

The table below shows the L3DAS22 Challenge rank for Task 2, including all the scores, for the sake of comparison.

Rank	Team Name	Precision ↑	Recall ↑	T2 Metric ↑
1	Lab9 DSP411	0.706	0.691	0.699
2	CQUPT AudioLab	0.600	0.584	0.592
3	FMSG	0.577	0.574	0.575
4	Team JLESS	0.587	0.561	0.574
5	SHRC of PKU	0.485	0.488	0.487
-	Baseline	0.423	0.289	0.343
6	DeepHearing	0.334	0.285	0.307
7	Anomaly	0.056	0.054	0.055



在本次挑战赛中，Lab9_DSP411团队提出一个具有新型数据增强方法的集成事件独立网络（Ensemble Event Independent Network）。由于不同模型的集成存在轨道排列问题，该集成模型也结合轨道输出格式、排列不变训练和多任务学习的软参数共享方法，旨在解决具有轨道输出格式的事件独立网络的集成模型问题。我们也采取了一种新型的数据增强方法，混合了随机生成的增强方法的数据处理技术，以提高模型的稳健性和不确定性估计。该提出的数据增强方法将一些增强操作随机采样和组合，以产生多样性的增强频谱图。我们提出的该轨道输出格式的集成模型系统，提高了模型的性能。该系统在L3DAS22 Task 2中获得了第一名，相较于第二名有显著优势。

A TRACK-WISE ENSEMBLE EVENT INDEPENDENT NETWORK FOR POLYPHONIC SOUND EVENT LOCALIZATION AND DETECTION

Jinbo Hu^{1,2}, Yin Cao³, Ming Wu¹, Qiuqiang Kong⁴, Feiran Yang¹, Mark D. Plumbley³, Jun Yang^{1,2}

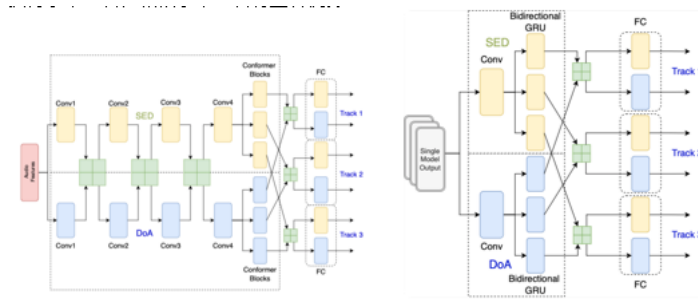
¹Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China {hujinbo, mingwu, feiran, jyang}@mail.ioa.ac.cn

²University of Chinese Academy of Sciences, Beijing, China

³Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, UK {yin.cao, m.plumbley}@surrey.ac.uk

⁴ByteDance Shanghai, China, kongqiuqiang@bytedance.com

该模型对应的论文：A Track-wise Ensemble Event Independent Network for Polyphonic Sound Event Localization and Detection也被ICASSP 2022接收，该论文将在今年5月 ICASSP 2022 大会的 Special Session与同行上进行分享。



下一篇：[研究人员揭示超声波对重油黏度和化学结构影响的机理 \(./202201/t20220125_6348689.html\)](http://202201/t20220125_6348689.html)



中国科学院
CHINESE ACADEMY OF SCIENCES

(<https://www.cas.cn/>)

旧版回顾 (<http://www.ioa.cas.cn/webold/>) | 网站地图 (<http://www.ioa.cas.cn/kjtd/wzdt/>) | 联系我们 (http://www.ioa.cas.cn/lxwm_2021/)

© 1996 - 2021 中国科学院声学研究所 版权所有 备案序号：京ICP备16057196号-1 (<http://beian.miit.gov.cn/>)

京公网安备110402500001号 (<http://www.beian.gov.cn/portal/registerSystemInfo>) 地址：北京市海淀区北四环西路21号中国科学院声学研究所

邮编：100190



(<http://www.ioa.cas.cn/wzjc/>)



(<https://bszs.conac.cn/sitename?method=show&id=0922261B20F17671E053022819AC9056>)

官方微信

