

論文 / 著書情報
Article / Book Information

題目(和文)	深層学習ベースの環境音セグメンテーション -音源定位、分離、クラス分類の統合-
Title(English)	Deep-Learning-based Environmental Sound Segmentation – Integration of Sound Source Localization, Separation, and Classification –
著者(和文)	周藤唯
Author(English)	Yui Sudo
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11924号, 授与年月日:2021年3月26日, 学位の種別:課程博士, 審査員:田中 正行,蜂屋 弘之,大山 真司,原 精一郎,宮崎 祐介
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11924号, Conferred date:2021/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第 号			学位申請者氏名		周藤 唯	
論文審査 審 査 員		氏 名	職 名		氏 名	職 名	
	主査	田中正行	准教授	審査員	宮崎 祐介	准教授	
	審査員	蜂屋 弘之	教授				
		大山 真司	准教授				
		原 精一郎	准教授				

論文審査の要旨（2000 字程度）

本論文は「Deep-Learning-Based Environmental Sound Segmentation – Integration of Sound Source Localization, Separation, and Classification –」（深層学習ベースの環境音セグメンテーション-音源定位、分離、クラス分類の統合-）と題し、英文全 6 章から構成されている。

第 1 章「Introduction(序論)」では、本論文の動機、目的、関連研究の課題、本論文の構成を述べている。これまで、それぞれ個別に評価されてきた環境理解の主要機能である音源定位、区間検出、音源分離、分類を統合的に扱うことの重要性を主張している。

第 2 章「Literature Review(関連研究)」では、音源定位、区間検出、音源分離、分類を統合的に扱う環境音セグメンテーションを定義し、関連研究の紹介・議論を行い、研究の位置付けを行っている。従来の統合手法では、課題 1) カスケード型統合による誤差蓄積による性能低下が生じる、課題 2) スペクトログラム上で音響イベントのサイズが異なると分類性能が劣化する、課題 3) 空間位相情報は周期性があるため、ニューラルネットワークでは効率的に扱うことができない、課題 4) 音源クラスと音源到来方向のような複数の変数を扱う場合、互いの関係に過学習してしまうといった問題があることを明らかにするとともに、本論文は、これらの課題を解決するため、複数タスクを統合的に扱うアプローチの検討、および技術の研究開発を行うことであると位置づけている。

第 3 章「Sound Event Aware Single-Channel Environmental Sound Segmentation（音響イベントに着目したシングルチャンネル環境音セグメンテーション）」では、課題 1)、2) に対し、音響イベント検出を導入した、音響イベントサイズのばらつきにロバストな **Mask U-Net** を提案している。シングルチャンネルのマイクロフォンを用いた環境音セグメンテーションタスクに対し、75 クラスの環境音を含むデータセットを作成し、評価を行っている。結果として、提案手法により、誤差の蓄積、音響イベントサイズのばらつきに対する性能劣化を抑えることができることを実証している。

第 4 章「Improvement of Sound Source Localization by using Quaternion for Sound Event Localization and Detection（音響イベントの定位と検出のためのクォータニオンを用いた音源定位の改善）」では、マルチチャンネルのマイクロフォンを用いた音響イベントの定位と検出タスクに焦点を当て、課題 3) について議論し、その解決手法を提案している。具体的には、音源信号のチャンネル間位相差や音源到来方向のような周期性を持つ空間情報に対し、クォータニオンから導かれる $\sin$, $\cos$ 関数ベースの特徴量を入力に用いることを提案、さらに、環境音データセットを用いた実験により、音源到来方向推定誤差の低減を確認し、提案手法の有効性を実証している。

第 5 章「Multichannel Environmental Sound Segmentation with Separately Trained Spectral and Spatial Features (スペクトル特徴と空間特徴を分離して学習したマルチチャンネル環境音セグメンテーション)」では、マルチチャンネルのマイクロフォンを用いた環境音セグメンテーションに焦点を当て、課題 4) の解決法検討を行っている。具体的には、音源到来方向ごとに音源分離を行うネットワークと音源クラスごとに音源分離を行うネットワークからなる統合的手法を提案している。また、75 クラスの環境音データセットを用いた評価実験を通じ、提案手法が、誤差の蓄積および音源クラスと音源到来方向の関係への過学習を緩和し、音源定位、分離、分類を効率的に統合できることを実証している。

第 6 章「Conclusion(結論)」では、個別の手法を統合する際に生じる上記 4 つの課題に対して、提案した環境音セグメンテーション手法に対する評価実験から得られた知見を整理し、学術的な貢献、および今後の課題について議論している。

以上を要するに、本論文は、環境理解の機能を個別に扱うのではなく、統合的に扱うことが肝要であるという観点から、環境音セグメンテーションを定義、これを実現するための独自手法を提案、その提案手法の有効性を実証している。こうした個別手法を統合的に扱うためのフレームワークの提案と実証は、環境理解分野の発展に工学上貢献するところが大きい。よって本論文は博士(工学)の学位論文として十分な価値があると認められる。

注意:「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。