数理・情報のフロンティア
2019 年度採択研究者

2019 年度
実績報告書

井上 中順

東京工業大学情報理工学院
助教

データ大統一に向けたマルチモーダル事前学習

§1. 研究成果の概要

本研究では、2020 年度までに「タスクを横断した音の事前学習」方法の確立を目指しています。具体的には、音声を解析して話者が誰かを認識する「話者認識」と、環境音を解析して何の音かを認識する「音響認識」の２つのタスクを同時に解くための方法を研究しています。

2019 年度(後期半年間)では、話者認識の評価実験フレームワーク作成をおこないました。これはインターネットから収集された、約 100 万の発話データからニューラルネットワークを学習し、話者の特徴を抽出するものです。上記の目標達成のための要素技術にあたります。また、フレームワーク作成と並行して、話者照合向けの新たな学習方法を設計しました。具体的には、ResNet18/34 と呼ばれる画像認識向けのネットワークを、話者照合向けに改良した上で、メタ学習セットを用いた新たな最適化手法を提案・実装しました。本成果に関しては、現在、論文をまとめている段階です。