

## 研究課題別事後評価結果

1. 研究課題名： データ大統一に向けたマルチモーダル事前学習

2. 個人研究者名

井上 中順（東京工業大学情報理工学院 准教授）

3. 事後評価結果

本研究では、音声認識と画像認識に共通して有効なニューラルネットワークモデルの構築を実現し、その事前学習手法について検討し、データの種類を横断できるような汎用的な学習手法の開発を目的としている。話者認識、環境音認識、顔認識、物体認識といった4つの具体的なタスクを研究対象とし、統合的な学習フレームワークの提案を行った。話者認識・環境音認識の研究では、大規模話者照合セットを用いた実験では世界最高レベルの性能を実現しており、その学術貢献は大きい。

音と画像を横断する学習手法についても映像データに基づき画像認識モデルから音認識モデルへの知識転移を行う手法を実現し、フラッグシップ国際会議 ICASSP で研究発表を行うなど、国際的に大きなインパクトを与えた。

今後、学習の安定性の向上と効率化などの課題を解決し、画像と音声に加えて言語情報をも横断できるような総合的かつ汎用的な学習手法の開発が期待される。