

22

音からものとかたちを認識

～音と画像のクロスメディア情報処理によるシーン理解～

どんな研究

音から物体や人物を認識する研究です。具体的には、マイクで収録した音だけを分析して、そこにある物体や人物などの種類や形状を含む画像認識結果を予測します。生活空間や公共空間での見守り・防犯等への応用に向けて研究開発を進めています。

どこが凄い

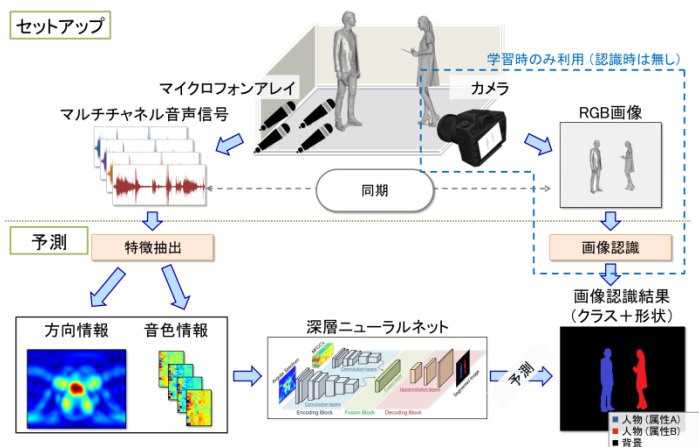
音のみから物体や人物の種類・形状を予測するには、音の方向と音色の時空間的な関係を捉える必要があります。本研究では、この複雑な関係を効率的にモデル化できる特徴統合層を備えた深層ニューラルネットワークを設計し、より多様な種類の物体の認識を可能にしました。

めざす未来

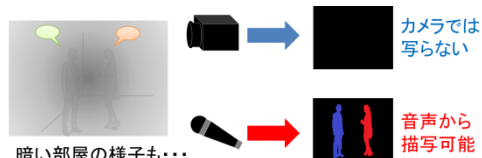
安心・安全な見守り・監視システムの実現が期待できます。本技術は、カメラが利用できないようなプライバシー性の高い家庭や公共空間への適用が可能です。また、音ならではの伝搬特性を活かし、カメラでは死角になってしまうような箇所の情報も予測できるようになる可能性があります。

“音から画像認識”するクロスメディア情景分析

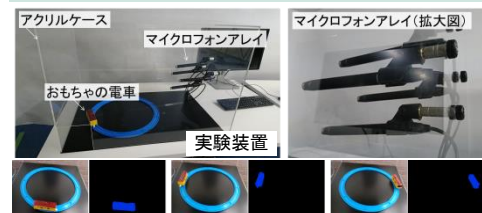
複数のマイクで収録した音の方向と音色を基に、音を発している物体や人物のクラス(種類)と形状を予測



カメラでは写らない・写したくない空間の様子を、音だけからでも描写可能に

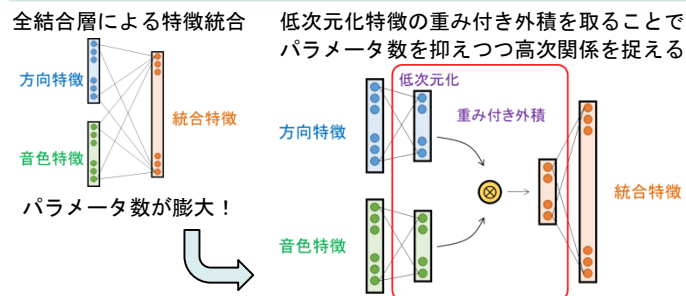


これまで、少数のクラスの物体や人物の認識が可能であることを確認

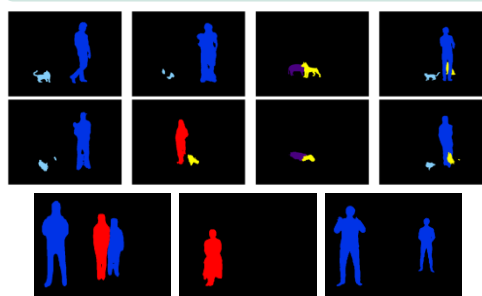


ここがポイント

方向と音色の関係を効率的に捉えることのできる特徴統合層を導入することにより、パラメータ数を抑えつつ精度を向上



より多様なクラスの物体や人物の認識、実音声からの認識を実現



関連文献

- [1] G. Irie, M. Ostrek, H. Wang, H. Kameoka, A. Kimura, T. Kawanishi, K. Kashino, "Seeing through sounds: predicting visual semantic segmentation results from multichannel audio signals," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019.
- [2] 王昊辰, 入江豪, 亀岡弘和, 木村昭悟, 平松薫, 柏野邦夫, "双線形特徴統合を用いた音からのセマンティックセグメンテーション," *第21回 画像の認識・理解シンポジウム (MIRU)*, 2018.

連絡先

入江 豪 (Go Irie) メディア情報研究部 メディア認識研究グループ
Email: cs-liaison-ml at hco.ntt.co.jp



Innovative R&D by NTT

オープンハウス 2019

Copyright © 2019 NTT. All Rights Reserved.