Keyword　マルチモーダル学習　規模推定　混雑予測

画像音声のマルチモーダル学習による規模推定・混雑予測に関する研究

情報科学研究科 情報科学専攻
データサイエンス・人工知能領域 博士前期課程
2025年3月修了

井手 伊織

主査 成 凱　副査 安部 恵介　澤田 直

研究背景

駅や観光地、商業施設などで起こりやすいオーバーツーリズムや雑踏事故の対策として、混雑が発生しやすい 場所に監視カメラを設置し混雑状況を監視したり、画像認識技術を用いて画像や動画から対象物を検出して人数や規模を推定 することで、混雑状況を自動的に判断する研究が進められている。しかし、密集度の高い場面に対応 できないことや、光の変化や遮蔽物の影響で精度が低下する問題が残されている。

研究目的

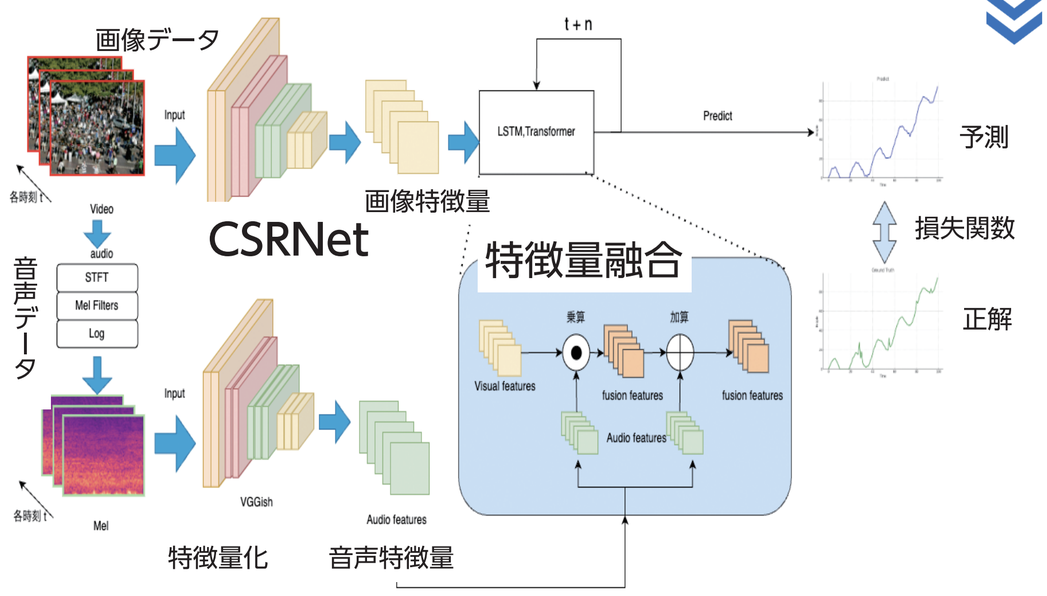
本研究では、音声情報を活用して空間内の混雑特性を補完し、音声特徴量を組み合わせ たマルチモーダル学習に基づく混雑予測モデルを提案する。画像特徴量として密度マップ、音声特徴量としてログメルスペクトログラムや事前学習済みモデル VGGish等を時系列予測モデルに組み込むことで混雑状況を予測可能なモデルを構築した。

研究概要

STEP 1. ビデオ撮影・画像音声データ収集

STEP 2. 画像アノテーション・正解データ作成

STEP 3. マルチモーダル学習による混雑予測



評価実験

モデル1 (CSRNet + LSTM)

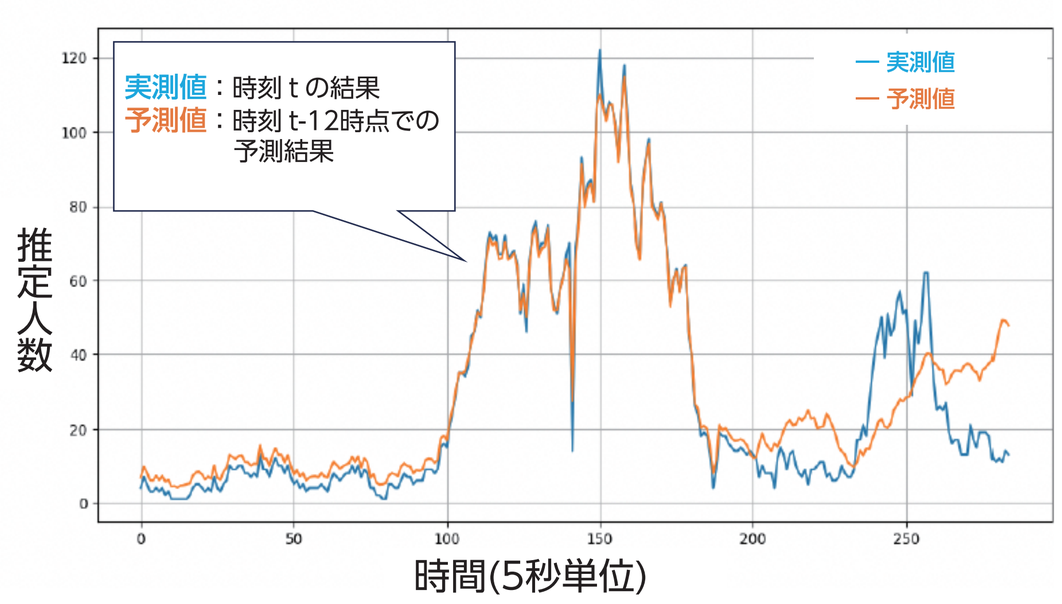


図1. 密度マップ・VGGish特徴量による予測

モデル2 (CSRNet +Transformer)

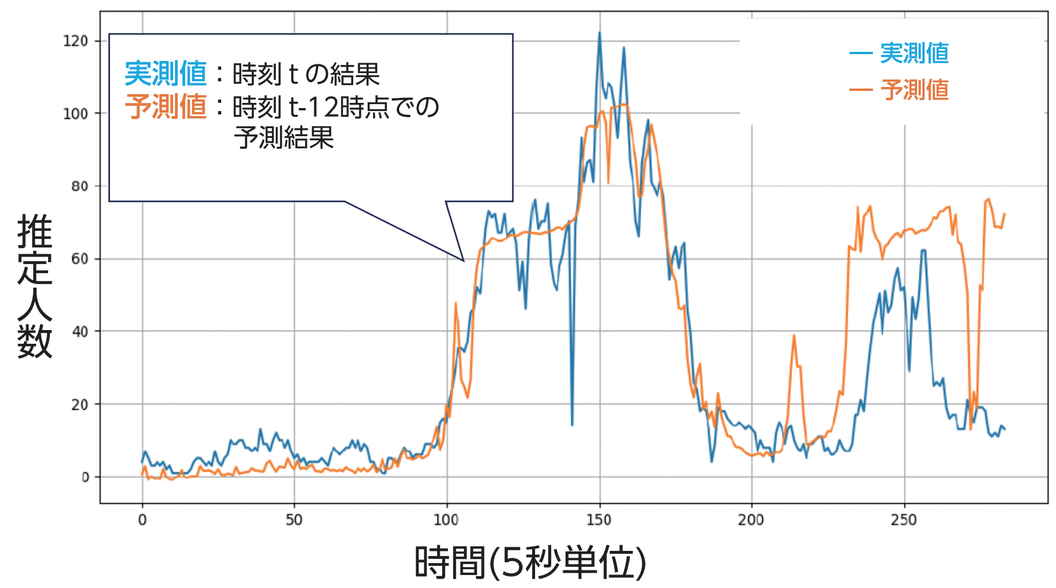


図2. 密度マップ・ログメルスペクトログラムによる予測

成果・まとめ

本研究では、オーバーツーリズムや雑踏事故の対策として駅や観光地などでの混雑状況の予測を目的として、画像と音声という異なる種類のデータを統合し、空間内の混雑特性を補完するマルチモーダル学習モデルを提案し実データによる評価実験を行った。実験結果として音声特徴量VGGishを用いることで予測精度が顕著に改善され、画像と音声を組み合わせたマルチモーダル学習が一定の有効性が示された。

指導教員コメント

本研究は、混雑状況の推定・混雑予測を目指して、従来画像のみでの群衆カウントに加え、光や遮蔽物等の影響を受けにくい音声データを活用することでより高精度な規模推定と混雑予測を可能にした。研究を進めるうえで、本学１号館フロアや正門前における動画撮影、画像アノテーション、予測モデルの構築と訓練等、多くの経験が得られて大きく成長したと思う。

成 凱