M-015

環境超音波に基づく屋内領域推定における SpecAugment によるデータ拡張の評価

Evaluation of data augmentation with SpecAugment for indoor localization based on environmental ultrasound

井上 貴裕† 梅澤 猛† 大澤 範高†
Takahiro Inoue Takeshi Umezawa Noritaka Osawa

1. はじめに

本研究では，環境内で動作する電気製品が発する超音波（環境超音波）を利用した屋内領域推定を行う．領域推定には，環境超音波から得たスペクトログラムデータを使って訓練した CNN（Convolutional Neural Network）モデルを用いる．この手法では，モデルの訓練に使用したデータと収集日が異なるデータに対する推定精度が低下する課題があることがわかった．そこで，本研究では時間および周波数方向でスペクトログラムにマスキングを施す SpecAugment によるデータ拡張を適用し，訓練データと異なる日のデータに対する汎化性能の向上を図った．データ拡張による汎化性能向上の効果を検証するため，訓練データとは異なる日に収集したデータを用いた領域推定実験を行い，その推定精度を評価した．

2. 関連研究

土谷らは複数の領域（部屋や廊下）で収集した環境超音波からスペクトログラムを生成し，CNN により分類モデルを構築することで，領域の推定を行った[1]．分類モデルによる部屋の推定精度は 97.5%で，環境超音波の解析を用いた屋内位置推定手法が高い精度を持つことが確認された．土谷らの研究では，訓練データとは異なる日のデータに対する汎化性能の評価が不十分であった．本研究では，特に訓練データとは異なる日に収集されたデータに対する領域推定の精度向上に焦点を当てた．

SpecAugment は，音声認識の分野において認識精度向上のために考案されたデータ拡張手法である[2]．スペクトログラムに対して，Time Warp（時間伸縮），Time Masking（時間マスキング），Frequency Masking（周波数マスキング）の 3 つの手法を適用する．時間マスキングと周波数マスキングにより，スペクトログラムの一部の情報を隠すことで，特定の時間や周波数の情報に依存しない汎化性能の高いモデルの学習が期待できる．本研究では，SpecAugment のこれらのマスキングを適用し，異なる日に収集されたデータに対する汎化性能の向上を図った．

3. 実験条件

3.1 録音

大学内の講義室や研究室，廊下など 10 領域を対象として環境超音波の収集を行った．録音はのべ 7 日（2023 年 10 月 26 日，11 月 9，16，30 日，12 月 7，14，21 日）実施した．録音時間は各領域 5 分間とした．ただし，モデルの訓練に用いる 10 月 26 日の録音については，5 分間の録音の後で 5 分のインターバルを設け，さらに 5 分間録音を行って，計 10 分間のデータを収集した．

超音波マイクには Dodotronic 社の Ultramic 384K EVO を使用した．実験ではマイクの最大サンプリング周波数 384kHz で録音を行った．超音波マイクは録音用のラップトップ PC に USB ケーブルを用いて接続し，机や台の上に置き静止した状態で音波を収集し，wav 形式で記録した．

3.2 スペクトログラム

実験では，特徴量として録音データを変換したスペクトログラムを用いる．スペクトログラムは音声信号に対して短時間フーリエ変換（short-time Fourier transform: STFT）を行い，得られた周波数スペクトルを水平方向に並べたものである．スペクトログラムでは，時間，周波数，強度の情報が 2 次元のヒートマップで表現される．

実験におけるスペクトログラムの生成手順について述べる．まず，収集した 5 分間の録音データをオーバーラップなしで 1 秒のデータに分割する．STFT の窓関数にハミング窓を使用し，フレーム長を 1,024 サンプル（0.0026 秒），フレームシフトを 512 サンプル（0.0013 秒）としてスペクトログラムを生成する．この際，人間の可聴域である 20kHz 以下は切り捨て，20kHz から 192kHz のスペクトログラムを使用する．この処理を経て，時間が 0.7 秒で周波数帯域が 20kHz から 192kHz を表す縦 512 画素，横 512 画素の画像として利用する．

3.3 CNN モデルの構造

本研究では，畳み込み層とプーリング層をそれぞれ 4 層ずつ持ち，全結合層が 2 層の構造のモデルを使用する．畳み込み層ではフィルタサイズを 7×7 とする，プーリング層では 3 つの最大プーリングの後にグローバル平均プーリングを行う．プーリング層のフィルタサイズは 3×3，ストライドは 2 とする．活性化関数には ReLU（Rectified Linear Unit）を用いる．正規化層ではバッチ正規化を行い，最後に全結合層を 2 層接続し，出力する．

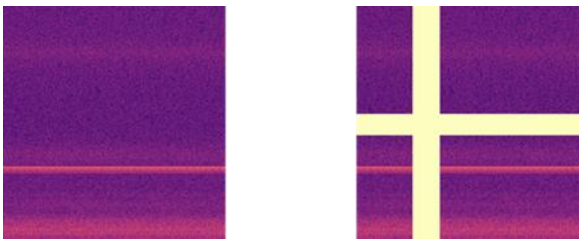
3.4 データ拡張

SpecAugment の時間マスキングと周波数マスキングを適用し，データ拡張を行った．図 1 に示すように，スペクトログラムの時間軸方向と周波数軸方向でそれぞれランダムな位置と範囲を選び，その範囲の，値を 0 とするマスキング処理を行った．このように，マスキングする区間の位置と範囲をランダムに変えることで，多様なパターンのデータ拡張を実現した．

時間マスキングでは，スペクトログラムの全時間長に対して最大 10%の長さの連続する区間をランダムに選び，その区間の値を 0 とした．同様に，周波数マスキングでは，

† 千葉大学 Chiba University

全周波数帯域に対して最大 10%の幅の連続する周波数帯をランダムに選び，その帯域の値を 0 とした．



(a) マスキング前　　(b) マスキング後

図 1　周波数マスキングと時間マスキングを施したスペクトログラム

4. 実験

10 月 26 日のデータを用いて学習を行い，その日とは異なる日に録音されたデータに対して推定精度の評価を行った．実験には表 1 に示すデータセットを利用した．10 月 26 日に録音されたデータから生成したスペクトログラムを重複するデータがなく，10 領域分のデータが同じ割合で含まれるように 5,000 件の訓練データと 1,000 件の検証データに分割して使用した．また，同一の日の録音データを基に，10 領域のデータを均等に含んだ 1,000 件のスペクトログラムでデータセットを構成し，互いに異なる 6 日分のデータセット 6 つを得た．推定精度の評価には，10 領域の分類における正解率を使用した．

モデルの訓練では，バッチサイズを 32 とし，損失関数にはクロスエントロピー誤差を用いた．最適化アルゴリズムには確率的勾配降下法を使用し，学習率を 0.001 に設定した．また，重み更新時のモメンタム係数は 0.9 とした． 10 エポック連続で検証データに対する損失が改善されなくなるまで訓練を行った．

表 1　各データセットの録音日とデータ数

データセット名	録音日	データ数
訓練データ	2023 年 10 月 26 日	5,000
検証データ	2023 年 10 月 26 日	1,000
テストデータ 1	2023 年 11 月　9 日	1,000
テストデータ 2	2023 年 11 月 16 日	1,000
テストデータ 3	2023 年 11 月 30 日	1,000
テストデータ 4	2023 年 12 月　7 日	1,000
テストデータ 5	2023 年 12 月 14 日	1,000
テストデータ 6	2023 年 12 月 21 日	1,000

最初に，データ拡張を施さずにモデルを訓練した場合の結果を示す．検証データに対しては 1.0 の精度が得られ，このことは，モデルが十分に訓練されていることを示している．各テストデータに対しては，0.890，0.740，0.980，0.994，0.795，0.966 の精度が得られた．6 つの精度の平均は 0.894 で標準偏差は 0.0967 である．

次に，データ拡張の効果を検証するため，訓練データに周波数マスキングと時間マスキングを施したスペクトログラム 5,000 件を追加した，合計 10,000 件の訓練データを使用してモデルを訓練した．CNN の訓練における条件はデータ拡張なしの場合と同様である．

データ拡張を施してモデルを訓練した場合の結果を示す．検証データに対して 1.0 の精度が得られた．各テストデータに対しては，0.986，0.846，0.989，0.990，0.861，0.949 の精度が得られた．6 つの精度の平均は 0.937 で標準偏差は 0.0607 である．

図 2 にデータ拡張を施さなかった場合と施した場合におけるテストデータに対する推定精度のグラフを示す．データ拡張を施さなかった場合と比較して，データ拡張を施した場合では，全てのテストデータセットに対して 0.846 以上の高い精度が維持されている．また，標準偏差が 0.0967 から 0.0607 となりデータセット間における推定精度のばらつきも低減されていることが確認できる．この結果は，時間周波数領域のマスキングによるデータ拡張が，モデルの汎化性能を向上させる効果をもつことを示唆している．

しかし，一部のデータに対してはデータ拡張を施した場合に精度が低下する場合も確認された．今後は，マスクする周波数帯や時間帯を固定して，その周波数帯や時間帯を変化させた場合の推定精度に対する影響を評価し，適切なマスキング条件について検証する．

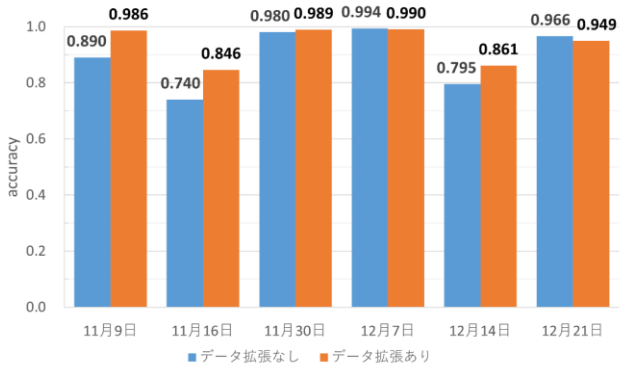


図 2　テストデータに対する推定精度

5. おわりに

本研究では，環境超音波に基づく屋内領域推定において，SpecAugment を利用したデータ拡張手法の効果を検証した．モデルの学習時とは異なる日に収集された 6 つのテストデータセットに対する平均推定精度が 89.4%から，データ拡張を用いた場合では 93.7%に向上することを確認した．この結果は，時間周波数領域のマスキングを利用したデータ拡張がモデルの汎化性能向上に寄与することを示唆している．

参考文献

[1] T. Tsuchiya, T. Umezawa and N. Osawa, "An Indoor Area Estimation Method Analyzing Spectrograms of Environmental Ultrasounds by Convolutional Neural Network," 2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS), Wuhan, China, 2018, pp. 1-7.

[2] Daniel S. Park, William C., Yu Z., Chung-Cheng C., Barret Z., Ekin D. Cubuk, and Quoc V. Le, “SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition,” Proc. Interspeech 2019, pp. 2613–2617, 2019.