

# スペクトログラム画像と U-Net を用いた 電車走行雑音の除去実験

林 堅<sup>†</sup> 市川 誉揮<sup>†</sup> 佐野 将太<sup>†</sup>  
川喜田 佑介<sup>†</sup> 宮崎 剛<sup>†</sup> 田中 博<sup>†</sup>  
<sup>†</sup> 神奈川工科大学 情報学部 情報工学科

## 1. はじめに

音のデータを画像データに変換して雑音除去や音の識別に適用することが行われつつある。本報告は U-Net を用いて雑音除去を行う検討において、従来の結果[1]より良好な結果が得られたので報告する。

## 2. 学習データの作成

本実験で使用する音声素材は「日本音響学会新聞記事読み上げ音声コーパス」の男性の音声データ(サンプリング周波数 16kHz)を用いた。この音声データに SN 比 0dB となるように、独自に取得した電車走行音を雑音として重畳する。そして、この重畳データを短時間フーリエ変換によってスペクトログラム画像に変換する。上記の一連の作業を Python プログラムにより自動化し、20 人分のデータセットを作成した。このデータセット内には訓練用画像 3530 枚と評価用画像 883 枚の学習画像とペアの教師画像(雑音重畳なし)を合わせて 8826 枚の画像が含まれている。

## 3. 実験手法と実験結果

本実験では深層学習のネットワークに U-Net を適用して学習モデルを作成した。図 1 の示すように、雑音を重畳した音声データのスペクトログラム画像に対して 9 回の畳み込みと 4 回のマックスプーリングを行い、画像を圧縮しながら、特徴量を抽出する。その後、4 回のアップサンプリングと 9 回の逆畳み込みを行い、画像を元のサイズに戻し出力する。出力画像と雑音を重畳する前の音声データである教師画像が一致するように U-Net を学習させる。

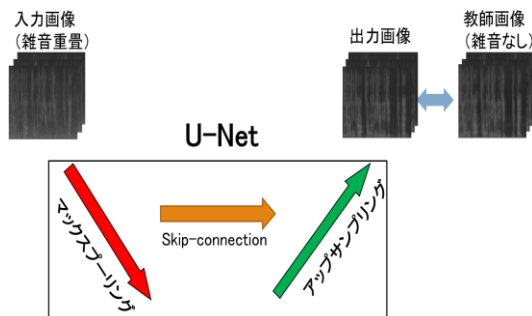


図 1 実験手法

今回は 20 人分のデータセットを学習し、モデルを作成した。学習パラメータを表 1 に示す。学習の収束状況を図 2 に示す。

表 1 学習パラメータ

エポック	バッチサイズ	最適化アルゴリズム	学習率	損失関数
200	5	Adam	1e-3	MSE

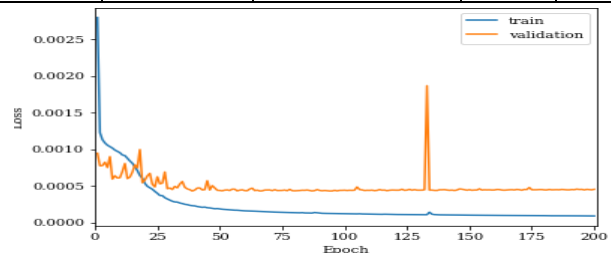


図 2 学習の収束状況

学習を終えたモデルに対して、新たに学習に用いていないテストデータでの雑音除去効果を評価する。雑音のない教師データのスペクトログラム画像を基準として、列車走行雑音を重畳した音声のスペクトログラム画像との相違度を図 3 に示す。今回の実験は差分二乗和で相違度を計算する。教師画像と雑音重畳後の画像との相違度が 0.0111 であった。今回作成したモデルによる雑音を除去した結果、相違度が 0.0027 まで大幅に減少した。相違度から良好な結果が得られたと考えられる。また、画像データから音への復元により音声を確認した結果も良好であった。

教師画像	雑音重畳後	除去結果
相違度	0.0111	0.0027

図 3 除去結果の評価

## 4. まとめと今後の課題

U-Net を用いて音声から雑音を除去する手法を適用し、十分な性能を得た。今後は SN 比が変化した場合での性能変化を調べ、ロバスト性の検証を行う。

### 参考文献

[1] 市川誉揮他, “スペクトログラム画像を対象とする U-Net を用いた電車走行音除去に関する検討,” 信学技報, SeMI2021-48, pp.61-65, 2021.