

環境騒音調査の音データを対象とした CBiGAN による異常音判別 Anomalous Sound Detection in Environmental Noise Measurement with CBiGAN

荒川永人[†] 荒川由人[†] 大谷紀子[‡]
Eito Arakawa Yuito Arakawa Noriko Otani

1. はじめに

環境基本法第 16 条では、人の健康の保護および生活環境の保全のうえで、最終的に、大気、水、土壌、騒音をどの程度に保つことを目標に施策を実施していくのかを示した環境基準が規定されている。環境騒音調査は騒音に係る環境基準の達成状況を評価するものである。環境騒音調査において対象とする騒音の範囲は、常態として存在する道路交通騒音であり、等価騒音レベルで評価することが原則となっている。等価騒音レベルは瞬間的な高レベル音に影響されることから、評価の妨げとなる異常音は除外する必要がある。環境省は評価マニュアル[1]において異常音の種類を定義している。除外対象とする異常音は、鳥の鳴き声などの平常ではない自然音、パトカーのサイレンなどの通常は発生しない人工音、測定員の咳払いなどの測定による付加的な音である。現行の異常音除外作業の具体例として、建設環境研究所で行われている処理手順を挙げる。建設環境研究所では、環境騒音調査において騒音計と IC レコーダを使いデータを記録している。異常音を除外するには、騒音計から取得した騒音の時刻歴データから異常音の疑いがある箇所を見つける。次に、IC レコーダで録音した異常音の疑いがある箇所の音を聞き、異常音か否かを判断する。以上のような異常音除外作業は多大な労力を要し、環境騒音調査における人間の負担が大きくなっている。各自治体における測定評価の現状についてのアンケート[2]によると、約半数の機関で有人での測定を行っており、環境騒音調査における測定員の負担は大きくなっている。また、無人で測定を行っている機関についても、録音した音や、等価騒音レベルに基づき除外の必要性を判断しており、現行の異常音除外作業は多大な労力を要する作業になっていることがわかる。先行研究では異常音除外作業における人の負担を軽減することを目的として、環境騒音調査の音データを対象とした Autoencoder に基づく異常音判別手法を提案したが、入出力の差分を異常度とする手法において出力の不鮮明さが課題だった。出力が不鮮明になると正常データの各ピクセルにおいて微細な再構成誤差が生じ、偽陽性の原因となる。本研究では生成モデルとして CBiGAN[3]を使用することで出力の不鮮明さを改善する。また、異常度算出過程においてパッチ分割と指数変換を導入することで微細な再構成誤差の蓄積による影響を減らす。

2. CBiGAN

CBiGAN は GAN を異常検知の用途で発展させた Efficient GAN[4]に基づいている。潜在変数から画像を生成する Generator と画像の真贋を判定する Discriminator で構

[†]東京都市大学大学院環境情報学研究所 Tokyo City University, Graduate School of Environmental and Information Studies

[‡]東京都市大学メディア情報学部 Tokyo City University, Faculty of Informatics

成される一般的な GAN に、画像を潜在変数にエンコードする Encoder を導入し、画像と潜在変数の生成を同時に学習するモデルである。CBiGAN では、正常画像のみを学習データとし、Encoder と Generator の学習、Discriminator の学習を敵対的に交互に実施することで、正常画像から潜在変数の生成、および潜在変数から正常画像の生成を同時に学習する。以上により、高精度な画像生成が可能な GAN を利用して再構成誤差に基づく異常検知が可能になる。

Generator を G 、Encoder を E 、Discriminator を D とし、テスト画像を X とするとき異常度 $A(X)$ は式(1)から算出され、 $A(X)$ の大きさに基づき異常か正常かを判別する。

$$A(X) = (1 - \lambda)\mathcal{L}_R(X) + \lambda\mathcal{L}_{fD}(X) \quad (1)$$

$$\mathcal{L}_R(X) = \|X - G(E(X))\|_1 \quad (2)$$

$$\mathcal{L}_{fD}(X) = \|f_D(X, E(X)) - f_D(G(E(X)), E(X))\|_1 \quad (3)$$

λ はドメイン固有のハイパーパラメータであり、 $f_D(\cdot)$ は Discriminator の中間層から抽出される特徴ベクトルである。テスト画像が異常画像の場合、正常画像だけを学習した Generator および Encoder は画像の再構成に失敗し、 $\mathcal{L}_R(X)$ が増加する。同時に、Discriminator は元画像 X と再構成 $G(E(X))$ から異なる特徴を抽出するため $\mathcal{L}_{fD}(X)$ が増加する。

3. 提案手法

提案手法では、環境騒音調査の音データを対象とし、CBiGAN を用いて異常音を判別する。異常度算出過程においてパッチ分割と指数変換を導入することで微細な再構成誤差の蓄積による影響を減らす。

3.1 前処理

前処理では音データをメルスペクトログラムに変換する。メルスペクトログラムは横軸を時間、縦軸を周波数、色を強度として音色の時間変化を画像で表現したものである。メルスペクトログラムは軸を持っているため、画像の反転などによるデータ拡張は適切ではない。したがって、メルスペクトログラムに変換する前の音データに対してノイズ付与、ピッチシフト、時間シフト、音量シフトを行い、音の聞こえ方自体を変化させることでデータを拡張した。

3.2 異常度

メルスペクトログラム上の細かいランダム模様を完璧に

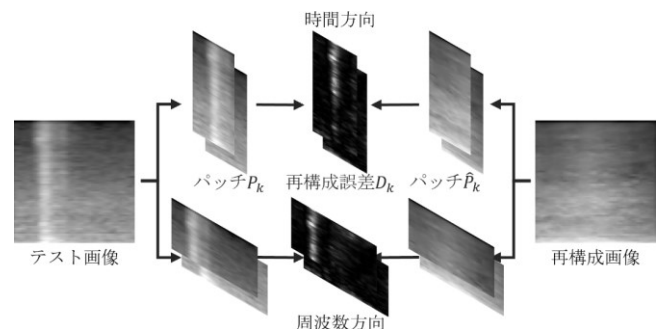


図 1 異常度算出手順

再構成することは非常に困難なタスクである。正常データにおける各ピクセルの微細な再構成誤差が蓄積すると、偽陽性の原因となる。本研究ではメルスペクトログラム上の異常領域を効果的に異常度に反映し、微細な再構成誤差の蓄積による誤判別を防ぐための異常度算出方法を提案する。異常度算出手順を図 1 および以下に示す。

1. テスト画像 X と再構成画像 $G(E(X))$ を時間方向および周波数方向に対して n 個ずつのパッチ $P_k, \hat{P}_k \in \mathbb{R}^{H_k \times W_k \times C_k}$ ($k = \{1, 2, \dots, 2n\}$) に分割する。
2. 式(4)より各パッチの絶対誤差 d を求め、式(5)より再構成誤差 D_k を求める。

$$d = |P_k - \hat{P}_k| \quad (4)$$

$$D_k = \sum_{H_k, W_k, C_k} (e^{4d} - 1) \quad (5)$$

3. 各方向の再構成誤差の最大値 $D_{\text{time}}, D_{\text{freq}}$ を求める。
4. 式(6)より $L'_R(\mathbf{x})$ を求める。
5. 式(7)より異常度 $A'(X)$ を求める。

$$L'_R(\mathbf{x}) = (1 - w)D_{\text{time}} + wD_{\text{freq}} \quad (6)$$

$$A'(X) = (1 - \lambda)L'_R(X) + \lambda L_{fD}(X) \quad (7)$$

異常音のメルスペクトログラムは図 1 のように、縦または横方向に特異なパターンがある。したがって、再構成誤差は、画像全体に対してではなく、時間方向および周波数方向のパッチごとに算出するほうが、全体の再構成誤差に対する異常領域の再構成誤差の割合が大きくなるため、異常領域の誤差を効果的に異常度に反映することができる。また、2 において、元画像と再構成画像の差分を指数関数によって変換することで、各ピクセルの微細な再構成誤差をより小さく、大きな再構成誤差をより大きくすることができる。

4. 評価実験

建設環境研究所が用意した音データから正常音を取り出しメルスペクトログラムに変換した 1068 枚の画像を学習データとし、別の日に測定された 1 日分の音データをテストデータとした。正常音とは環境騒音調査において評価の対象となっている道路交通騒音である。

本実験で使用したモデルについては、CBiGAN[3]における Encoder, Generator, Discriminator の残差ブロックを計算効率化を目的としてボトルネック型の残差ブロックに変更した。また、異常度算出手順において、式(4)の重み λ は 0.4 とした。各方向のパッチ分割数 n は 2 とし、式(5)における周波数方向の重み w は 0.3 とした。

実験では、Autoencoder, CBiGAN, 異常度 $A'(X)$ を使用した提案手法、以上 3 つの手法の有効性を PR-AUC を用いて評価した。なお、建設環境研究所が用意した正解ラベルには、異常音と正常音が交互に混在する時間に対してまとめて異常ラベルが付与されており、一般的な再現率が使用できない。本実験では異常ラベルが付与された区間のうち、異常と判断されたデータを含む区間の割合を再現率として PR-AUC を算出した。テストデータにおける PR-AUC を表 1 に示す。また、異常音と正常音それぞれに対して出力された異常度の分布を図 2 に示す。

表 1 PR-AUC

	Autoencoder	CBiGAN	提案手法
PR-AUC	0.802	0.810	0.860

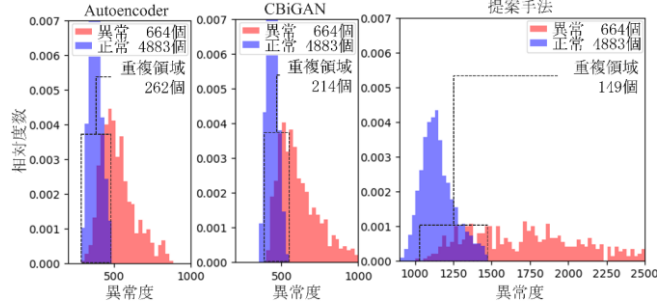


図 2 異常度の分布

5. 考察

長時間の音データに含まれるすべての異常音に対して正確に異常ラベルを付与することには限界がある。したがって、異常音の開始・終了時刻を正確に把握し網羅的に異常音を判別できているかについて厳密に評価するのは難しい。しかし、表 1 より汎用的な判別性能は提案手法が最も高いことがわかる。また、図 2 に示すように異常音と正常音、それぞれに対する異常度分布の重複領域に含まれるデータ数は提案手法が最も小さいことから、提案手法によりメルスペクトログラム上の異常領域を効果的に異常度に反映できていることがわかる。

一方で、提案手法がベースとしている CBiGAN は学習の安定性、処理速度の面で課題がある。学習データの増加やハイパーパラメータの変更によってモデルの性能が大幅に変化すると同時に学習に膨大な時間を要する。したがって、さまざまな地点で収集された学習データに対して適切なネットワーク構成やハイパーパラメータを逐一設定しモデルを構築するのは非常に困難である。

今後の展望としては、CBiGAN と他の手法とのアンサンブルが挙げられる。Autoencoder に基づく手法は安定性や判別精度が比較的高く、処理速度も比較的速いが、トラックやバイクの走行音の誤判別が多い。CBiGAN は小規模なデータセットに対しては比較的安定して学習できることから、学習データをトラックやバイクの走行音に限定することで特定の走行音を高精度に判別できるモデルを構築することができる。したがって、CBiGAN と Autoencoder をアンサンブルすることで Autoencoder の欠点を CBiGAN によって補うことができ、精度の向上が期待される。

謝辞

本研究を進めるにあたり、多くのデータを提供していただきました株式会社建設環境研究所に深謝いたします。

参考文献

- [1] 環境省, “騒音に係る環境基準の評価マニュアル 道路に面する地域編” (2015).
- [2] 全国環境協会議会関東甲信静支部 騒音振動専門部会研究連絡会, “「環境基準の評価から除外すべき音」の取扱いについて (道路に面する地域)”, 全国環境研会誌, Vol. 27, No. 4, pp. 51-55 (2002).
- [3] F. Carrara, G. Amato, L. Brombin, F. Falchi and C. Gennaro, “Combining GANs and AutoEncoders for efficient anomaly detection”, 2020 25th International Conference on Pattern Recognition (ICPR), pp. 3939-3946 (2021).
- [4] H. Zenati, C. S. Foo, B. Lecouat, G. Manek and V. R. Chandrasekhar, “Efficient GAN-Based Anomaly Detection”, arXiv:1802.06222 (2018).