実環境下におけるデータの取得制約を考慮した異常音検知

畔柳 伊吹

論文要旨

本論文は製造設備や社会インフラの学習ベースの異常音検知を対象とし，現場導入で避けられない二視点から体系的な枠組みを提案する．第一は環境起因の変動である．異常音検知モデルの学習時と運用時で背景雑音やマイク配置，機械の動作設定が変わると正常音の入力音響特徴の分布がずれ，異常スコアと閾値が不安定になる．第二はデータ可用性である．異常ラベルや機械の属性ラベルの取得度合いで，現実に選べる異常音検知モデルの学習戦略は大きく変わる．本論文はこの二軸で課題を構造化し，現場で機能する異常音検知モデルの設計原則と運用指針を提示する．

本研究では異常音検知モデルの設計の中心として直列法を採用する．前段では背景雑音や伝達系の影響を抑え，機械挙動に敏感な特徴量抽出器を学習する．後段では得られた埋め込み上で異常検出器を用いて距離や尤度に基づく異常スコアを計算し，研究指標と単一閾値運用をつなぐ．この分離により環境変動への頑健性と更新の容易さの両立を図り，停止時間や再配布の制約下でも段階的な適応を可能にする．

データ可用性は，異常ラベルの有無と属性ラベルの有無で四条件に整理する．異常ラベル無，属性ラベル有という標準的条件を基準に置き，少量の異常データが得られる条件と，ラベルが一切ない導入初期の条件を主対象とし，異常ラベルが得られるものの属性ラベルが欠落する条件も補助的に扱う．どのラベルがどこまで入手できるかを，学習戦略と運用設計に直結させる枠組みを構築する．

異常ラベルが得られない標準的条件に対し，擬似異常を用いる学習を改善する手法として Serial-Outlier Exposure（Serial-OE）を提案する．前段の特徴量抽出器を強化し，

過学習を抑えつつ検知性能とデータ効率の向上を目指す．国際的な異常音検知コンペティションのデータセットである Detection and Classification of Acoustic Scenes and Events（DCASE）2020 Task 2 において，全マシンタイプにおける AUC と partial AUC（p=0.1）の平均値で 93.5 を達成し，既存最良法（Noisy-ArcMix）に対して約 +1.6pt の改善を示した．さらに運用中に少量の実異常音が得られる場合には，それらを擬似異常集合へ逐次統合して再学習することで追加の性能向上が得られ，異常サンプル 1 件でも平均値が +2.2pt 改善することを示した．

異常ラベルも属性ラベルも一切使えない条件では，外部データから対象機械に近いサンプルを選別して取り込み，データ間の関係から擬似ラベルを生成し，反復再学習する手続きを提案する．これにより，環境要因ではなく機械挙動に根ざした埋め込みを前段で形成し，後段の距離や尤度に基づく検知の実現を目指す．DCASE2022–2024 Task 2 の未ラベル条件において，全データセット平均 AUC で 68.9 を達成し，未ラベル条件におけるベースライン手法の性能 63.6 を +5.3pt 上回った．また，属性ラベルを用いたベースライン手法との性能差を 9.4pt から 4.3pt へ縮小し，ラベル欠如による性能低下を 5.1pt 緩和した．

本論文の貢献は三点に集約される．第一に，環境軸とデータ軸の二視点から問題空間を体系的に整理し，四条件の分類により現場のデータ取得制約を研究設計に直接マッピングする枠組みを確立する点．第二に，少量の異常データが得られる場合とラベルを使えない場合に対し，直列法を共通骨格とする具体的な学習戦略を提案し，Serial-OE と外部データ選別・擬似ラベル反復に基づく自己教師あり学習手続きによりデータ効率と汎化の両立を図る点．第三に，データ可用性ごとに推奨戦略と運用上の留意点を体系化し，前段と後段の分離更新，現場で得た異常の段階的取り込み，外部データの選別的利用を，実務的判断材料として提示する点である．環境の揺らぎとラベル欠如という現実の制約下でも破綻しにくい設計と，更新の負担を抑えた運用手順を研究から実装まで一貫して示すことが，本研究の中核的な意図である．

Abstract

This dissertation addresses learning-based anomalous sound detection for manufacturing equipment and social infrastructure systems and proposes a systematic framework from two perspectives that are unavoidable in real-world deployment. The first perspective is environment-induced variation. When background noise, microphone placement, or machine operating settings differ between training and operation, the distribution of acoustic features of normal sounds shifts, making anomaly scores and thresholds unstable. The second perspective is data availability. The degree to which anomaly labels and machine attribute labels can be obtained greatly changes the learning strategies that can realistically be chosen for anomalous sound detection models. By structuring the problem along these two axes, this dissertation presents design principles and operational guidelines for anomalous sound detection models that function in practice.

This study adopts a serial approach as the core design of anomalous sound detection models. In the first stage, a feature extractor is learned that suppresses the influence of background noise and transmission characteristics and is sensitive to machine behavior. In the second stage, an anomaly measure is computed on the learned representation based on distance and likelihood, bridging research metrics and single-threshold operation. This separation aims to balance robustness to environmental variation with ease of updating, enabling stepwise adaptation even under constraints such as downtime

and redeployment.

Data availability is organized into four conditions based on the presence or absence of anomaly labels and attribute labels. Taking as a standard condition the case where anomaly labels are unavailable but attribute labels are available, this dissertation mainly focuses on (i) conditions where a small number of anomaly samples can be obtained and (ii) conditions at the initial deployment stage where no labels are available at all, while also treating as a supplementary case the condition where anomaly labels are available but attribute labels are missing. This work builds a framework that directly links which labels can be obtained and to what extent to the learning strategy and operational design.

For the standard condition in which anomaly labels cannot be obtained, we propose Serial-Outlier Exposure (Serial-OE) as a method to improve learning with pseudo-anomalies. By strengthening the first-stage feature extractor, the method aims to improve detection performance and data efficiency while suppressing overfitting. On the dataset of the international anomalous sound detection competition, Detection and Classification of Acoustic Scenes and Events (DCASE) 2020 Task 2, it achieved 93.5 in the mean of AUC and partial AUC ($p = 0.1$) across all machine types, showing an improvement of approximately +1.6 points over the previous best method (Noisy-ArcMix). Furthermore, when a small number of real anomalous sounds become available during operation, we show that additional performance gains can be obtained by incrementally integrating them into the pseudo-anomaly set and retraining, and that even a single anomaly sample improves the mean performance by +2.2 points.

For the initial deployment stage where no labels can be used, we propose a procedure that selects and incorporates samples from external data that are close to the target machine, generates pseudo-labels from relationships among the data, and iteratively

retrains. This enables the first stage to form embeddings grounded in machine behavior rather than environmental factors, and aims to realize detection in the second stage based on distance and likelihood. Under the unlabeled condition of DCASE2022—2024 Task 2, we achieved 68.9 in the mean AUC across all datasets, outperforming the baseline method under the unlabeled condition (63.6) by +5.3 points. We also reduced the performance gap relative to the baseline method using attribute labels from 9.4 points to 4.3 points, mitigating the performance degradation due to missing labels by 5.1 points.

The contributions of this dissertation are summarized in three points. First, it systematically organizes the problem space from the two perspectives of the environment axis and the data axis, and establishes a framework that maps real-world data acquisition constraints directly to research design through a four-condition classification. Second, for cases where a small number of anomaly samples can be obtained and cases where labels cannot be used, it proposes concrete learning strategies with the serial approach as a common backbone, and achieves both data efficiency and generalization through Serial-OE and a self-supervised learning procedure based on external-data selection and iterative pseudo-label refinement. Third, it systematizes recommended strategies and operational considerations for each level of data availability, and presents, as practical decision-making material, the separated updating of the first and second stages, the stepwise incorporation of anomalies obtained in the field, and the selective use of external data.

The central intent of this study is to consistently demonstrate, from research through implementation, a design that is unlikely to break down under real-world constraints such as environmental fluctuations and missing labels, along with operational procedures that reduce the burden of updates.

目 次

図 目 次

表 目 次

第1章

序論

1.1　背景

1.1.1　実環境における異常音検知の特徴と制約

製造設備や社会インフラの安定稼働を支える状態監視において，異常音検知は非接触・低侵襲で導入でき，稼働中の機械が発する微細な変化を音として捉えられる有力な手段である．また，センサの設置自由度が高く，暗所や機械内部など視認が困難な環境にも適用可能である [4], [5], [6].

一方で，学習ベースの異常音検知モデルの開発における最大の制約は異常事象の希少性にある．安全・品質の観点から意図的に故障を再現してデータを収集することは難しく，得られる異常データは少量かつ偏りやすい．この制約のもとで，学習ベースの異常音検知システムを構築する際にどのようなデータを用いて学習するか，すなわち学習設定の設計が性能を大きく左右する．

本章の狙いは，観測モデルの前提を明確化し，環境起因の変動とデータ側の制約という二つの軸から課題を整理し，本論文の研究焦点を位置付けることにある．以下ではまず学習ベースの異常音検知における学習設定を整理し，つづいて問題設定と前提を定め，課題を環境軸とデータ軸に分類し，代表的手法の説明へと進む．その上で環

境軸の詳細と採用する設計原則を導出し，データ軸の論点を述べる．

1.1.2 学習設定の整理と本研究の立場

異常音検知の学習設定を整理するため，本稿では正常ラベルを $y = 0$，異常ラベルを $y = 1$ とする二値ラベルを用いる．学習に提供されるデータとラベル可用性により，主に次の三つの設定に大別できる [7], [8]：

- 教師あり：正常音と異常音の双方に対し，それぞれの状態を示すラベルが付与されたデータセットを用いる．代表的異常サンプルを直接参照できる一方，異常な状態の網羅は困難であり，クラス不均衡や過適合に留意が要る．
- 半教師あり：正常音ラベル $y = 0$ が付与されたデータのみを用いて学習する．データ収集の現実性が高く，未知異常の広いカバーを狙える反面，訓練データとして利用可能な正常データの網羅性が不十分だと，訓練データと評価データの条件差によって正常データを異常として過検知する問題が生じうる．
- 教師なし：ラベルが付与されていない正常音と異常音の混合データセットから学習する．大規模データ活用に適するが，正常集合の汚染率に関する仮定やロバスト設計が必要である．本論文では直接の対象外とする．

本論文ではもっとも一般的な半教師あり設定を基盤として議論する．

1.2 問題設定

1.2.1 観測モデルと前提

異常音検知は，時刻 t における収録音 $x(t)$ に対して，ラベル $y \in \{0, 1\}$（0=正常，1=異常）を推定する二値分類問題である．なお，ラベル y は収録された音響クリップ

全体に対して単一のラベルが付与されるものとする．収録音 $x(t)$ は

$$x(t) = h(t) * \{ s_c(t) + n(t) \} \tag{1.1}$$

と表す．ここで，$c \in \mathcal{C}$ は機械型や運転設定などの動作属性ラベル，$s_c(t)$ は属性 c のもとで対象機械が発する動作音，$n(t)$ は背景雑音，$h(t)$ はマイク・設置位置・音場を含む伝達系で，$*$ は畳み込みである．

本研究では以下の前提を置く：

- A1（半教師あり学習）：学習データは原則として正常ラベル $y = 0$ を持つサンプルを主体とする．
- A2（加法雑音・独立）：背景雑音 $n(t)$ は動作音 $s_c(t)$ と独立な加法項であり，任意の属性 c に対して $s_c(t) \perp n(t)$ とみなす．

非定常性や同期雑音が強い場合には式 (1.1) は近似となる．以降の議論は上記前提の範囲で進める．

1.2.2 環境軸とデータ軸の課題

異常音検知の課題は，環境軸の変動とデータ軸の制約に大別される [9], [10], [11], [12]．環境軸では，学習条件と運用条件の不一致（ドメインシフト）が異常スコア分布や単一閾値運用の安定性を損なう．データ軸では，目的変数 y，属性ラベル c，メタ情報 $(n(t), h(t))$ の可用性が学習設計に影響を与える．こうした課題を論じる前提として，まず代表的手法を整理し，その視点から環境軸を比較し，本研究で採用する設計方針の直列法を導出する．

1.3 異常音検知手法の二分類：生成モデルと識別モデル

異常音検知の代表的手法は，大きく生成モデルと識別モデルに分かれる [13]．

1.3.1　生成モデル

生成モデルは，観測信号から抽出された入力音響特徴 $\boldsymbol{x}$ に対して，正常分布 $p(\boldsymbol{x} \mid y = 0)$ を推定し，逸脱度を異常スコアとして用いる．具体例として，オートエンコーダによる再構成誤差を利用するもの [14], [15], [16], [17], [18]，Long-Short Term Memory ネットワーク [19]，WaveNet [20]，生成的敵対ネットワーク [21], [22] などがある．また，ガウス混合モデル（GMM）の尤度 [23], [24], [25] や正規化フロー [26], [27], [28] を利用する手法も提案されている．これらの手法は正常サンプル間の連続的な変化を柔軟に表現できる一方，収録音全体をモデル化するため，正常・異常の判断には本質的に関係ないものの，訓練時には観測されなかった未観測の背景ノイズ $n(t)$ や伝達系 $h(t)$ の変化によって尤度が体系的に低下し，異常の過検知に傾きやすいと報告されている [27], [29], [30]．

1.3.2　識別モデル

識別モデルは，動作音 $s_c(t)$ の属性 c に基づく補助タスクを通じて，パラメータ ϕ を持つ特徴量抽出器 f_ϕ を学習し，入力 $\boldsymbol{x}$ を低次元の埋め込み $\boldsymbol{z} = f_\phi(\boldsymbol{x})$ へと写像する．最大クラス確率の負値や近傍距離などで異常スコアを算出する．具体例として，データ拡張を用いた機種分類に基づく手法 [31]，メタデータ分類に基づく手法 [1], [32], [33], [34], [35], [36], [37]，対照学習に基づく手法 [38], [39], [40], [41]，正常データと擬似異常データを独自に定義して分類する手法 [42], [43], [44], [45], [46], [47], [48] などがある．識別モデルは動作音の属性ラベルに着目した特徴量の獲得を目的とするため，背景ノイズ $n(t)$ や伝達系 $h(t)$ を不要因子として抑圧しやすく，環境変動に対して生成モデルより相対的に頑健である [35]．ただし識別モデルを用いる場合は得られた埋め込みに対して異常スコアの設計が必要である [49], [50], [51]．

1.4 環境軸の課題と設計の採用

本節では，式 (1.1) を前提に，環境起因の変動に対する生成・識別の挙動差を整理する．環境軸の課題とは，訓練時と評価時における背景雑音，伝達系，動作設定の違いによって生じる変化（ドメインシフト）がモデルの挙動に影響を与えることを指す．ドメインシフトによって異常スコア分布や最適な閾値が変わるため対応策が不可欠である [9], [10], [11]．本研究では，ドメインシフトを不要因子 $(n(t), h(t))$ とタスク関連因子 c に分離して扱う設計を採る．

1.4.1 背景雑音と伝達系：不要因子への対処

式 (1.1) のもとで観測信号は

$$x(t) = h(t) * \{\, s_c(t) + n(t) \,\}$$

で与えられる．ここで $(n(t), h(t))$ はタスクと無関係な不要因子であり，c は機械タイプ・運転状態などのタスク関連因子である．識別モデルは，c を推定する補助タスクに基づく学習を通じて，$(n(t), h(t))$ に起因する変動の影響を相対的に抑圧した表現を獲得することを狙う [35].

具体的にはパラメータ ϕ を持つ特徴量抽出器 f_ϕ により

$$\boldsymbol{z} = f_\phi(\boldsymbol{x}) \in \mathbb{R}^D, \qquad \hat{c} = g_\psi(\boldsymbol{z}), \quad (\phi, \psi) = \arg\min_{\phi,\psi} \ \mathbb{E}\big[\ell(g_\psi(f_\phi(\boldsymbol{x})), c)\big] \tag{1.2}$$

となるように学習する．ここで，D は埋め込みベクトルの次元数，g_ψ はパラメータ ψ を持つ識別器であり，入力 $\boldsymbol{x}$ は観測信号 $x(t)$ から抽出された入力音響特徴とする．式 (1.2) により得られた埋め込み $\boldsymbol{z}$ は c の識別に十分な情報を保持するよう学習される．このとき，c の識別に直接寄与しない背景雑音やチャネル等の変動は，表現上で相対的に抑圧されることが期待される．

観測空間で直接異常スコアを算出する場合，その値は c に加えて $(n(t), h(t))$ にも強く依存し，学習条件と運用条件の乖離に伴って正常データのスコア分散が増大し得る．生成モデルではこの分散増大が性能低下に直結しやすい一方，識別モデルでは前段の学習により不要因子の影響を低減した表現を得た上で，後段での異常検知を行うことで運用時の安定性を向上できる [35]．したがって，本研究では中間表現である埋め込み $\boldsymbol{z}$ 上で異常スコアを定義・算出することで，$(n(t), h(t))$ の影響を前段の特徴量抽出過程で抑圧する方針を採る．

なお，雑音やチャネルの影響が大きい場合，埋め込み空間におけるクラス間マージンが圧縮し，異常スコアの分布，特に正常データの分布の幅が狭まることがある．そのため，スコア設計と分布整合が重要となる [10]．

1.4.2 動作側の変化：タスク関連因子への対処

生成モデルと識別モデルはいずれも追加データによる追学習で新たな属性 c に適応しうるが [52], [53]，運用上は再学習なし，もしくは最小限のサンプルでの適応が望ましい．本論文では，前段で特徴量抽出器 f_ϕ により $(n(t), h(t))$ の変化に頑健な埋め込み $\boldsymbol{z}$ を確保し，後段では $\boldsymbol{z}$ 空間上で異常検出器 A を用いて近傍距離や尤度などを計算して，属性 c に伴う運転条件や機種間の個体差のような連続的な変化を受容する．この設計は，(i) 不要因子の抑制，(ii) 埋め込みにおけるクラス間マージンの確保，(iii) 柔軟な後段スコアリングによる過剰適合と過少適合のバランス，を段階的に満たすことで汎化と運用安定性を両立できる [11]．さらに，k-nearest neighbor (kNN) [54]，local outlier factor (LOF) [55]，GMM [23] などの異常検出器を用いることによって閾値調整の簡素化や，前段の特徴量抽出器を温存したまま後段の異常検出器の交換・再学習が可能になるといった運用上の利点もある [56]．

以上より，本研究の環境軸に対する基本方針は，前段で識別モデルを用いて c に紐づく埋め込み $\boldsymbol{z} = f_\phi(\boldsymbol{x})$ を学習することで $(n(t), h(t))$ の影響を抑圧し，後段で $\boldsymbol{z}$ 上

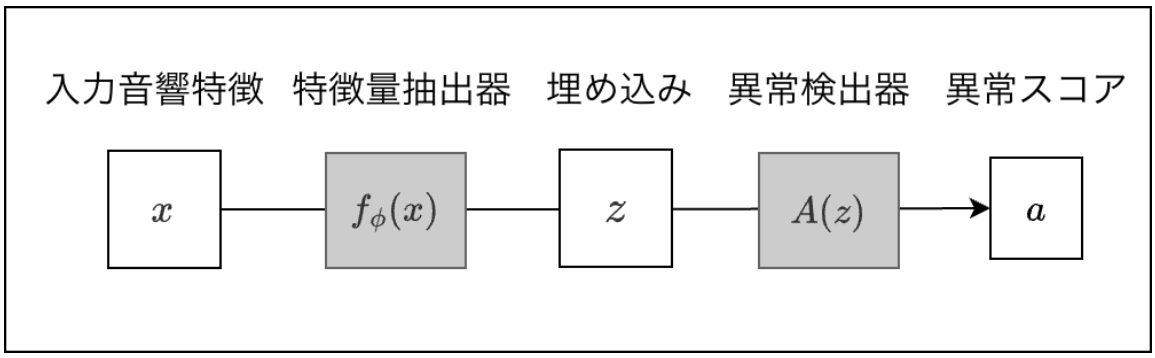


図 1.1: 直列法に基づく異常音検知手法の概要図

で異常スコア a を設計・整合させる直列法を採用する．直列法の概要を図 1.1 に示す.

1.5 データ軸の課題と本研究の焦点

データ軸の課題を整理するために，学習時に外部から与えられる情報を，目的変数そのもの（異常ラベル y），正常構造を記述する補助ラベル（属性 c），環境メタ情報（$(n(t), h(t))$）に分けて考える．最終判断は y によって定義されるため，y は最も直接的な情報であるが希少である．そこで，属性ラベル c を用いて埋め込みを表現することで正常な状態を定義し，そこから外れたサンプルを異常として扱うことができる．一方，$(n(t), h(t))$ は前処理や拡張，正規化に有用なメタデータだが，これらのみを用いて目的変数 y を直接定義することはできない.

以上より，教師データとして外部から与えられる情報は y と c に集約される．学習条件は異常ラベルの有無と属性ラベルの有無の二つの軸で整理でき，表 1.1 に示す四象限に分類される.

表 1.1 では，半教師あり学習を基盤に四象限に分類した．A は正常データのみを用いる標準的な半教師あり設定であり，本研究の基準として用いる．B は異常ラベルと属性ラベルが利用可能な設定であり，異常データを手に入れた際にその活用が容易であることが求められる．C は完全にラベルがない設定であり，ラベルがない状況でも埋

表 1.1: データ軸における学習条件の分類

	c 利用可（属性ラベルあり）	c 利用不可（属性ラベルなし）
y 利用不可 （異常ラベルなし）	A：半教師あり（第 3 章・第 4 章）	C：完全ラベルなし（第 4 章）
y 利用可 （異常ラベルあり）	B：異常あり（第 3 章）	D：異常のみ（第 3 章）

め込みを獲得するためのモデルを訓練することが求められる．D は異常ラベルのみが存在し属性ラベルが欠落する設定に相当する．本論文では B と C を主たる対象とし，D は B に対する小規模な追加実験として補助的に扱う．

1.6　本研究の貢献

本論文の貢献は，以下の三つの階層に整理される．

貢献 1: 問題空間の体系的整理と設計原則の確立

産業応用における異常音検知の課題を，環境軸とデータ軸という二つの視点から体系的に整理する．環境軸では，式 (1.1) に基づき背景雑音 $n(t)$ や伝達系 $h(t)$ といった不要因子を識別的に抑圧する必要性を明確にする．データ軸では，異常ラベル y と属性ラベル c の可用性に基づき，学習条件を A–D の四象限 (表 1.1) に分類する．この整理により，現場で直面するデータ取得制約を研究設計に直接マッピング可能な枠組みを提供する．

さらに，環境変動に頑健な表現を前段で獲得し，その表現上で異常スコアを計算する後段を組み合わせる「直列法」を，全章を通じた共通の設計原則として採用する．この二段構成により，前段と後段を独立に更新・管理できる運用上の利点が期待される．

貢献 2: データ制約下での具体的学習戦略の提案

表 1.1 の分類において，従来研究で十分に扱われていなかった B 領域 (異常データあり) と C 領域 (完全ラベルなし) に対し，直列法の枠組みを拡張した具体的な手法を提案する．

B 領域では，第 3 章において，擬似異常の枠組みに実異常データを統合する学習戦略を提示する．正常と擬似異常の二値分類に加え，属性ラベルに基づく識別をマルチタスク学習することで，少量の異常データでも過学習を抑えつつ性能改善を図る．また，属性ラベル欠落時 (D 領域) への適用可能性も検討する．

C 領域では，第 4 章において，異常ラベルも属性ラベルもない条件に対して，以下の三段階による自己教師あり学習戦略を提案する：(i) 外部データからの擬似異常サンプルの選別的導入，(ii) データ間関係に基づく擬似ラベル付与，(iii) 反復的再学習による段階的な性能向上．これらにより，ラベル取得が困難な導入初期段階でも異常検知システムの構築を可能にすることを目指す．

貢献 3: 実運用を見据えた設計指針の提示

データ可用性（表 1.1 の A–D）に応じた推奨学習戦略と運用上の留意点を体系化する．具体的には，以下の実務的判断材料を提供する：(i) どのデータ収集を優先すべきか，(ii) モデル更新をどの段階で行うべきか，(iii) 閾値をどのように設計するか．

さらに，前段と後段の分離更新，段階的な異常データ取り込み，外部データの選別的利用など，ダウンタイム制約下でも性能を維持・向上可能な運用手順の確立を目指す．

1.7　本論文の構成

第 2 章では，異常音検知の基礎として，特徴量抽出，識別学習，異常スコア推定に加え，ドメインシフトへの対応，評価指標（AUC / pAUC）と単一閾値運用の詳細を整理する．第 3 章では，半教師あり学習を基盤としつつ異常データを有効に活用できる手法を提案し，提案手法の検知性能を異常データの有無の双方で評価する．第 4 章では，属性ラベルがない枠組みにおいて活用可能な手法を提案し，各提案による効果を検証する．第 5 章では設計指針の提示と今後の展望を述べる．

第2章

関連研究

2.1　はじめに

異常音検知モデルを構築する際には，入力信号の整備から異常スコアの算定までに多岐にわたる要素技術が関わる．本章ではこれらの技術を整理し，後続の章で提案する手法の理解を助ける．まず，前処理や入力音響特徴の抽出方法を概観し，続いて識別モデルを用いた埋め込みの獲得を目的とした損失設計を紹介する．次に，異常スコア推定の代表的な方法を説明し，ドメインシフトへの対応としてドメイン適応とドメイン汎化の違いと具体的な対処法を整理する．最後に，実運用で用いる閾値の設計と評価指標との関係を述べる．

2.2　入力音響特徴の設計

異常音検知では，入力信号の前処理と入力音響特徴の抽出方法が後段の性能に大きな影響を与える．本節では，入力信号に対してどのような前処理が施され，どのような入力音響特徴が用いられているのかを概観する．

2.2.1　前処理

異なる収録条件でも一貫した入力を得るためには，録音区間の長さやサンプリング周波数を揃え，振幅スケールを正規化する必要がある．これらの情報を統一することによって，後段のモデルが環境条件の影響を受けにくい特徴を学習しやすくなる．多くの場合，モデルへの入力サイズを固定するために，収録音を固定長の窓を設けてチャンクとして分割する [42]．収録音が固定長に満たない場合はゼロパディングまたは固定長まで音源を繰り返すようにパディングを行い [57]，固定長よりも長い場合は，一つのチャンクが固定長になるように重複分割して後段でスコアを集約する場合が多い [48]．

2.2.2　入力音響特徴の抽出

入力信号を時周波数表現に変換する際，サンプリング周波数は異常音が含まれる周波数帯域を十分に含む値に統一し，高域や低域の不要成分は必要に応じてフィルタリングする [11], [13]．フィルタバンクを機械ごとに調整することで性能改善する報告もあるが [58]，異常データが希少な半教師あり設定では機械ごとのフィルタバンクの最適化は困難である．高周波成分が低周波成分よりも重要であるという報告もあるため，機械ごとに調整をするのではなく一律に高周波成分を強調する場合もある [59]．時周波数表現には短時間フーリエ変換またはログメルスペクトログラムが広く採用されている [60], [61]．近年は，単一の時周波数表現だけでなく，複数の表現を用いることでニューラルネットワークが自動的に有用な特徴を活用できるようにした方法 [3] や，メルスペクトログラムに加えて一次元畳み込みニューラルネットワークを用いて波形から直接抽出する方法 [33] も提案されている．

2.2.3 正規化

入力音響特徴に対してスケールを揃えるために正規化を行う場合がある．波形自体を単位分散にスケーリングする方法 [46], [48] や，スペクトログラムの周波数軸方向に対して正常サンプルの平均を引き標準偏差で割る方法 [31], [62] が用いられる．機種によっては雑音抑圧のため Teager–Kaiser エネルギー演算子を用いる方法 [63] も提案されている．一方で，正規化手順を明示しない研究もあるため，入力音響特徴の標準化は必須ではないことを示している [32], [64]．なお，ニューラルネットワークを用いるほぼすべての手法において，中間表現に対してバッチ正規化 [65] を適用し，学習の安定化と高速化を図っているため，バッチ正規化の活用は重要と考えられる．

2.3 識別モデルを用いた特徴量抽出

異常音検知では，学習時に属性ラベルを用いた補助的なクラス分類で埋め込み空間を形成し，推論時には正常分布からの逸脱度をスコア化する枠組みが広く採用されている [11], [13]．このように識別モデルを用いて異常データと関係のない補助ラベルを分類する手法は Outlier Exposure（OE）とも呼ばれ，正常音のみで学習する設定においても有効である [9]．以下では OE の枠組みにおける代表的な埋め込み空間の訓練方法を紹介する．表 2.1 に，本節で扱う特徴量抽出器の代表的設計を，必要情報と目的の観点から整理する．

2.3.1 属性ラベルを用いた多クラス分類

すべての属性ラベルを用いて多クラス分類する方法では，クラス間分散の最大化とクラス内分散の最小化を同時に高めるため，角度マージン系の損失関数が用いられる場合が多い．代表例として ArcFace [2] や AdaCos [66] があり，異常音検知では属性

表 2.1: 識別モデルを用いた特徴量抽出器の代表的設計と必要情報

手法カテゴリ	必要情報	目的関数の要点	利点	欠点
多クラス分類	属性ラベル c	c 識別に有用な埋め込みを学習し，クラス間マージンを確保する	学習が安定しやすい	属性設計に依存
距離学習	属性ラベル／擬似ラベル	同クラス近接と異クラス分離を直接促し，頑健な距離構造を形成する	多峰性に適応しやすい	ペア設計と計算量に依存
Outlier Exposure による二値分類	正常データ, 擬似異常の定義	正常と擬似異常を識別させ，正常領域の境界を形成する	異常不要で導入が容易	擬似異常設計に依存
事前学習モデルの活用	事前学習モデル，少量適応データ	汎用表現を保持しつつ識別タスクで整形し，埋め込みを形成する	少量でも性能が出やすい	過学習と更新範囲の管理

ラベルを用いた補助分類タスクと組み合わせて性能の高い埋め込みを得る報告が多数ある [37]．さらに，クラス内に複数のモード（サブクラスタ）が存在する状況を明示的に扱うために，Sub-Cluster AdaCos (SCAdaCos) が提案され，機種ごとの多峰性を反映したプロトタイプ学習と分布推定を可能にしている [1]．より近年には，角度マージンとサブスペース射影を組み合わせ，補助分類タスクの難易度や表現スケールに適応する AdaProj が提案され，その有効性が示されている [67]．以上のような角度マージン系損失は，補助分類に基づく埋め込み形成において広く採用されている．

2.3.2　距離学習

属性ラベルや擬似ラベルを用いた距離学習も有効である．トリプレット損失やコントラスト損失により，クラス内分散の最小化とクラス間距離の最大化を行い，頑健な埋め込みを獲得する手法も提案されている [3], [39], [40], [41], [64]．例えば，同一機器の異なる設定間の表現を洗練するための対照学習 [40]，生成モデルとコントラスト学

習を結合した GeCo [39]，角度マージンと Mixup [68] を統合してコンパクト性とマージンを同時に高める Noisy-ArcMix [41] などが提案されている．

2.3.3 擬似異常データの活用

異常音検知では正常音のみで学習する設定が基本であるため，外部音や加工音を擬似異常として用い，正常との二値分類で特徴量抽出器を鍛える枠組みも用いられている [42], [43], [44], [45], [46], [47], [48], [69]．これらの手法は，正常と擬似異常の定義方法や二値分類器の設計により精度に影響を受けるため設計が重要である．正常と擬似異常の定義方法は大きく分類すると 2 通りある．一つ目は，動作設定ごとに一つの設定に対して一つの正常を定義し，それ以外の設定を全て擬似異常として定義する方法である．この方法はモデル間の埋め込みの分散が大きく単一閾値による運用が困難な点，正常と異常の違いと比較して正常と擬似異常の識別が容易すぎる場合に，埋め込みにて正常と異常の違いが表現されないという課題がある [42], [46]．正常データをより精緻に表現するために，機種ごとに正常と擬似異常を定義する方法が提案されており，より高い性能を示している [48], [70], [71]．擬似異常データの活用を前提としているため，実際の異常データが得られた場合にもその活用が比較的容易である点は特徴の一つである．

2.3.4 事前学習モデルの活用

近年は，音響領域の PANNs [72]，BEATs [73]，EAT [74]，OpenL3 [75] などの事前学習済み特徴量抽出器をファインチューニングや低ランク適応（LoRA）[76] により軽量に適応し，その埋め込みを kNN [54] でスコア化する実装が報告されている [3], [77], [78], [79]．特に異常音検知のようにドメインシフトの影響が大きい状況下では，LoRA により少量データでも属性分類器を適応させる構成の方が，過学習を抑えつつ事前学

表 2.2: 埋め込み空間における異常検出器の比較

区分	代表例	スコアの定義	利点	注意点
非パラメトリック	kNN	参照集合 $\mathcal{Z}_0$ との局所距離・密度に基づく	多峰性に強い，実装が容易	参照点数と距離尺度に依存
パラメトリック	Mahalanobis, GMM	分布仮定の下で密度・尤度（または距離）を推定しスコア化	小規模参照でも扱える場合がある	推定が不安定になり得る

習の表現力を温存できると報告されている [80]．この流れは，補助分類で埋め込みを整え正常分布からの逸脱をバックエンドで測る方法と整合しており，事前学習モデルの活用は今後も発展が期待される [12]．なお，BEATs や EAT は自己教師あり学習により事前学習された汎用オーディオ特徴量抽出器であり，事前学習自体は生成的・自己回帰的な側面や離散トークン予測に基づく表現学習である．一方，異常音検知での活用は LoRA による属性分類などの識別タスクによって微調整し埋め込みを形成する点にある．本論文では，これらの事前学習モデルを識別モデルを用いた特徴量抽出の枠組みに含めて記述する．

2.4　埋め込み空間における異常検出器

埋め込み空間での異常スコアは，特徴量抽出器で得た埋め込み z に対して，正常集合との局所距離，密度，尤度などを異常検出器で計算することで得られる．本節では，埋め込み空間におけるパラメトリックな尤度・密度推定に基づく手法（例：マハラノビス距離，GMM）から，分布仮定を置かない kNN による距離ベース手法に至る流れを整理する．表 2.2 に，本節で扱う異常検出器をスコアの定義と仮定の観点からまとめる．

2.4.1 非パラメトリック手法：距離・密度に基づく推定

距離ベースの異常検出器では，参照集合 $\mathcal{Z}_0$ に対して，埋め込み $\boldsymbol{z}$ の kNN 平均距離を異常スコアとして用いる．ここで $\mathcal{Z}_0 = \{\boldsymbol{z}_j\}_{j=1}^{N}$ は正常データから得た参照埋め込み集合である．kNN は（i）分布仮定を置かない非パラメトリック，（ii）多峰性やクラス混在に強い局所評価，（iii）高次元でも不安定な共分散推定を要さない，（iv）コサイン距離に整えた埋め込みと相性が良いといった理由から近年の報告で採用例が増えている [3], [78]．実装面でも，FAISS [81], [82] などの近似最近傍探索を用いることによって大規模データでも実用的であり，閾値設計やスコア整合との組み合わせが容易である．

2.4.2 パラメトリック手法：分布仮定に基づく推定

パラメトリックな尤度・密度推定に基づく異常検出器として，単峰の正規分布を仮定したマハラノビス距離 [83] や，多峰性を扱うために埋め込み空間に GMM を適合させ，負の対数尤度 $-\log p(\boldsymbol{z})$ を異常スコアとする方法 [23] が古典的かつ有効な方法である．しかし，ドメインシフトが発生し，そのターゲットとなる正常データが少数であり，複数の設定条件による分布の多峰性が同時に存在する設定では，（a）共分散推定の不安定性，（b）正規分布仮定の誤り，（c）GMM の初期値・正則化依存性が過検知や尤度のスコア範囲のばらつきを招き，単一閾値による評価を困難にする場合がある．そこで近年は前段で識別的に整えた埋め込みに対して，仮定の少ない kNN を適用する構成が選択されやすい．

2.4.3 推論時のスコア集約

推論では固定長チャンクごとのスコアが得られるため，クリップ単位の意思決定へ接続する後処理として集約が必要となる．集約方法を検討する際は，短時間に突発的に発生する異常には最大値，連続的に発生する異常には平均値を用いるなど，観測対

象に応じた適切な対応が必要である．補完的な戦略として，各スコアのうち異常スコアの高い上位 $p\,\%$ の平均を取る方法も提案されている [48].

2.5　ドメイン適応とドメイン汎化

本節では，トレーニング時に十分な正常データが得られたソース条件と，運用時に生じる別条件のターゲット条件を区別し，条件差（ドメインシフト）に頑健な設計を整理する．ここでいう条件差とは，マイク種類や設置位置の違い，背景騒音の種類や大きさ，機械の運転設定の相違などであり，これらは異常有無そのものより大きな信号変動をもたらすことが多い．近年の異常音検知タスク，特に国際的なコンペティションである DCASE (Detection and Classification of Acoustic Scenes and Events) Task 2 では，評価時に条件ラベルが与えられず，一つの閾値で判定する設定が一般的である [10], [12].

2.5.1　ドメイン適応（Domain Adaptation）

ドメイン適応は運用環境へ迅速に追従させるための選択であり，ダウンタイムや計算資源が限られる現場では軽量な手段から試すのが合理的である．異常音検知では次のような実装が報告されている．

まず，特徴量抽出器は維持しつつ後段の異常スコア計算を条件別に切り替えるバックエンド適応がある [47], [84]．例えば選択的マハラノビス距離は，入力と再構成の残差に対して条件別（ソースとターゲット）に推定した共分散を用い，残差の距離のうち小さい方を異常スコアとすることで条件差に起因するスケール不一致を低減する [84]．この方式はニューラルネットの再学習を伴わず，ターゲット正常の少数標本から統計量だけ更新できるため，停止時間やデータ共有の制約が厳しい運用に適する．

次に，ニューラルネット側を軽量に適応させる例として，条件依存統計を担う正規

化層だけをターゲット正常で更新する方法や，ドメインアライメント層をネットワークに挿入して各層の中間表現分布を条件間で近づける方策がある．後者は AutoDIAL に代表され，各層の条件差を内部で緩和してから異常スコアを計算する [63], [85].

さらに積極的な適応では，ターゲット正常の数ショットを用いた微調整やメタ学習で特徴量抽出器自体を迅速にターゲットへ追従させる．少数データで過学習しないようプロトタイプ損失やメタ更新を組み合わせる設計が提案されている [86], [87]．一方，こうした適応後にソース条件の性能が劣化する場合があるため，ソース・ターゲット両条件を併用した正則化やバックエンド側の条件別推定と併用して劣化を抑える工夫が必要となる [88].

2.5.2 ドメイン汎化（Domain Generalization）

ドメイン汎化は，特定のターゲットに合わせて再学習せずとも未知条件に破綻しない表現とスコア計算を設計する方法である．異常音検知では以下の四つの観点が有効である．

(1) マルチドメイン対応： 学習時のミニバッチで条件の出現頻度を釣り合わせる，あるいは SMOTE [89] や Mixup を用いて少数条件を合成的に増やすことで，学習が特定条件に偏らないようにする [25], [90]．また条件別の小規模モデルと条件推定器を併用し，出力を統合する二段構成も提案されている [90]．これらは条件ごとに強いが，条件数だけ構成が増えやすく維持管理コストが高い.

(2) ドメイン不変表現： 自己教師ありの対照学習や統計整合を活用して中間表現の条件依存成分を抑える方法が提案されている．DG-Mix は自己教師あり事前学習の目的関数に条件間・仮想条件間の差を抑える項を加え，Mixup で生成した仮想条件も含めて頑健性を高める [91]．また，条件ごとの共分散の差を最小化する正則化により表現の条件差を抑える報告もある [92]．前処理としてピッチシフトや時間伸縮，帯域フィルタ等の拡張で条件差の揺らぎを広げ，学習時からドメインの多様性を経験させるこ

とも効果的である [92].

(3) 特徴量の分離：環境由来の揺らぎと，劣化や故障兆候などの機械本体の挙動を同じ特徴量で扱うと，「環境が違うだけなのに異常と判定する」誤警報につながる．このため，両者を明示的に分けて学習する設計が提案されている．たとえば，収録環境や運転モードなどドメインシフト要因に関わる補助タスクを同時に学習し，条件に依らず安定して使える軸と，条件で変わる軸を切り分けるマルチタスク学習 [93]，および勾配反転層やフォーカル損失で「ドメイン差では説明できないズレ」だけを強調する手法 [94] がある．より踏み込む例では，正規化フローなどを用いて観測音を「物理パラメータに対応する潜在」と「残差」に分け，運転条件で説明できない成分を異常候補として扱う [95] ほか，階層メタデータに基づき機械タイプ / 運転モードといった粒度別のプロトタイプ距離を学習し，どの差異を許容すべき通常変動とみなすかを段階的に管理する手法もある [96]．これらの手法は，どの揺らぎを「許容すべき正常」とみなすかをモデル側に明示的に埋め込む点が重要であり，十分なメタ情報がある現場では，単一閾値での運用安定化に特に有効となる．

(4) 異常スコア計算の工夫： 表現がある程度ロバストになっても，スコア計算が条件ごとにスケールや分布形状の差を増幅すると単一閾値で評価できないという課題が生じる．これに対して，学習後に条件ごとに再構成残差の共分散を推定し，条件別の距離を計算したうえで近いほうを最終スコアとする選択的マハラノビス [84] は，条件間のスケール差を抑えるアプローチである．別のアプローチでは，テスト点から参照点への距離をその参照点周辺の近傍との平均距離でスケーリングし，各参照点で得た比の最小値をスコアとすることで，条件間でばらつくスコア範囲を圧縮する [97]．さらに，ソースとターゲットのドメインごとに kNN 距離を標準化 (z-normalization) し，それぞれの正規化距離のうち近い方を採用するドメイン正規化＋最小統合 [98] では，スコア分布のスケール差とドメインずれを同時に抑えられる．これらの方法は特徴量抽出器の再訓練を要さず既存システムの後段だけを更新できる利点がある．

総じて，現場での更新コストや停止時間の制約が厳しい場合は，バックエンド適応とスコア計算の工夫を優先的に適用し，余力があれば自己教師あり事前学習や属性を用いた多タスク学習によって表現自体を堅牢化するのが費用対効果の高い順序である．

2.6 閾値設計と評価指標の整合

モデルによって得られた異常スコアを実運用で迷いなく使うためには，異常ラベルに依存しない方法で正常と異常を区別するための閾値を定めるとともに，研究段階の評価指標と運用時の目標とを整合させることが重要である．各年の DCASE Task 2 では評価時に条件ラベルが与えられず単一閾値での判定が求められるため [10], [12]，ここでも半教師ありの前提と矛盾しない設計を採る．

2.6.1 正常データのみを用いた閾値の設計

閾値を設定する際には狙いたい誤警報率（FPR）α を定める．次に，スコア整合後の正常スコアだけを集め，その分布の上側 $100(1-\alpha)$ 分位を閾値 τ とする．直感的には，正常であるにも関わらず異常と誤判定される割合が α を超えないよう，正常スコアの上位 α をちょうど切り落とす位置に線を引くという考え方である．

他の方法として正常スコア分布へガンマ分布など一峰の連続分布を当てはめ，推定した形状・尺度から所望の分位を計算する方法も用いられる [10]．この近似は厳密さよりも再現性と保守性を重視した設計であり，現場でのパラメータ共有や監査にも向く．

2.6.2 評価指標との整合

本節では, 研究段階の評価指標と運用時の目標との整合を述べる．研究段階では AUC や pAUC のような閾値非依存指標が手法間の系統比較に適している．その理由として，

閾値に基づく評価を採用した場合，機械によっては同じシステムを用いた場合でも大きな性能差が出る可能性があるためである．したがって，研究段階の系統比較には閾値非依存の AUC / pAUC を用いる．一方，運用で重視されるのは定めた FPR でどれだけ検出できるかである．したがって実務では評価は次の二本立てにするのが分かりやすい：(i) 分位閾値 τ を実際に適用したときの FPR や F1 の実測値を併記する，(ii) そのうえで全体的な識別能力の比較には AUC / pAUC を用いる．近年では AUC に変わる評価指標として F1-EV は良好な閾値を推定できる可能性をスコア化する指標として実務上有用であると報告されている [99]．本研究では他手法との比較のため AUC と pAUC を主要指標として採用するが，運用時は観測対象の性質に応じて閾値を固定した評価や F1-EV を用いた評価も併用することが重要と考えられる．

2.7　まとめ

本章では，半教師あり異常音検知における関連研究を要素技術の観点から体系的に整理した．入力前処理では固定長化と時周波数表現，正規化やデータ拡張が，環境条件に起因する変動の影響を抑えた表現とスコア算定を行う上で重要であることを述べた．識別モデルを用いた埋め込み学習では角度マージン損失や距離学習，擬似異常（OE），事前学習モデルの活用を通じて，属性識別に有用で，かつ条件依存成分が過度に支配しない表現を獲得する方法を紹介した．異常スコア推定では局所距離や密度に基づく手法を中心に整理した．続いてドメインシフトに対する適応・汎化のアプローチを整理し，単一閾値運用に接続するための分位ベースの閾値設計と評価指標の整合を示した．これらの基盤技術は第 3 章で提案する少量異常データの活用や第 4 章で提案するラベルなし設定へのアプローチを支えるものであり，次章以降ではこれらを踏まえて新たな手法とその実験結果を示す．

第3章
Serial-OE：異常データを学習に活用可能とするOutlier Exposureと直列法に基づく異常音検知

3.1　はじめに

本章では，運用段階で断片的に得られる異常データを，どの段階で活用すべきかを整理し，その要件に適合する異常音検知手法を提案する [100]．この課題は表 1.1 における B・D の領域を対象とする．異常データの活用の方針は大きく二つに分かれる．一つ目は，学習済みの特徴量抽出器を固定したまま，異常検出器のみを更新する方針である．具体的には，新たに得られた少数の正常・異常サンプルの埋め込みを参照集合に逐次追加し，距離や類似度に基づくスコア計算を即時に改善する．二つ目は，異常データを特徴量抽出器の訓練に組み込み，特徴量抽出器と異常検出器の双方を再適応させる方針である．後者では，異常データが示す周波数成分の出現や周期性の崩れなど，構造的な差異を特徴量抽出器が表現できるように調整することで，未知異常に対する識別余裕を広げることが期待できる．

どちらが優れるかは運用の前提によって異なる．停止時間を許容できず再配布が難しい現場では，後段のみの更新が現実的である．定期停止や再配布が可能で検知力そのものの底上げを重視する場面では，異常データを訓練側へ取り込む再学習が有効となる．本章の関連研究では，先に後段のみを更新する報告例を整理し，その後に実異常を用いた再学習・再適応の枠組みをまとめる．

3.2　関連研究

3.2.1　異常データを用いた後段のみの更新

学習済みの特徴量抽出器を保持し，後段の異常スコア計算のみを更新する枠組みが報告されている．代表的には，運用中に不確実と判断された入力に対して人手で正常および異常ラベルを付与し，得られた埋め込みを参照バンクに逐次追加して，最近傍距離や類似度に基づくスコアをオンラインに更新する方式である．誤警報として確認されたサンプルを正常参照として蓄積し，境界の安定化を図る設計も提案されている [101]．この方式はシステム停止やモデル配布を伴わずに性能改善を反映できるという運用上の利点がある一方で，特徴量抽出器そのものは固定のため，未知異常に対する最終的な検出能力は元の特徴量抽出の識別性能に依存する．

対比として，少数の実異常を直接参照して感度を高めるSNIPER [102] やSPIDER-Net [103] といったfew / one-shotな手法も，前段の特徴量抽出を固定したまま参照集合と閾値を更新するという点で広義の後段のみ更新に属する．ただし，これらは異常種が追加されるたびに参照集合の管理やスコアの再較正が生じやすく，タイプ横断で単一閾値運用を目指す場合には運用負荷が増大しやすい．一方，[101] のように参照バンクを逐次拡張しつつ決定関数自体は保つ設計では，閾値ドリフトは相対的に小さいが，やはり前段の特徴量抽出を固定する限り，性能の上限はその表現力に依存する．

3.2.2 実異常を用いた再学習

異常音検知で異常データを訓練側に取り込む研究は多くない．既存の枠組みとしては, 正常と擬似異常の二値学習を基盤に, 異常データを擬似異常の一部として統合して学習する設計がある [42], [46], [104]．Deep Double-Centroids Semi-supervised Anomaly Detection（DDCSAD）損失と Binary Cross Entropy（BCE）をマルチタスク学習する手法は，正常機械の特定の動作設定を「正常」，同じ機械の異なる動作設定や他機械種の正常音を「擬似異常」として定義し，特徴量抽出器を二値分類と距離学習を用いて訓練する．異常データを活用する場合は，その擬似異常側へ異常データを加えることで，その特徴を反映する [42]．この手法自体は直列法を用いてはいないものの，識別モデルによって特徴量を抽出する方法は直列構成と親和的であり，前段で環境ノイズの影響を抑えつつ動作音に着目する埋め込みを獲得し，後段で正常機械の特定の動作設定ごとの局所密度に基づきスコア化するという分担により，運用差によるスケールの揺らぎを抑えやすいという利点がある [1].

一方で動作設定ごとにモデルを定義する必要があるため，モデルの管理コストが高くそれぞれのモデルごとの異常スコアの分布がモデルごとに異なるため単一閾値による運用が困難という課題がある [42]．この課題を回避するために，識別モデルは機種ごともしくは全ての機種で統一のモデルを用いることが望ましい [70], [71].

補足として, 画像分野では, 外部異常を負例として取り込む手法 [105], [106], CutMix によって局所不連続な画像を合成して学習を強化する手法 [107], [108]，異常データや外部データを損失に直接組み込む半教師あり手法 [107], [109] が体系化され，既知異常の検出力と未知異常への頑健性の両立が報告されている．しかし，これらの手法は学習時に画像特有の前提に依存するため，そのまま音響へ適用することは難しい．第一に，画像では画像同士の重ね合わせが非自然事象であるため Mixup や CutMix といったデータ拡張に対して異常性を担保しやすいが，音では線形加法が通常運転でも頻出するため，単純な混合は正常混音と区別しにくい [107], [108]．第二に，画像では欠陥

が空間内の局所に現れパッチごとに独立に扱えるという前提の下で，正常境界の外側を学習させる設計が多い [107], [109], [110]．これに対して音響では，時間と周波数にわたる連続性や長期の周期性が異常の主要手掛かりであり，短時間のパッチ切り貼りや位置固定のマスクは，その連続性と周期構造を断ち切ることになるので異常な特徴を返って検知しにくくする可能性がある．これらの違いから，本研究ではこの音データ特有の要件に適合する形で，前段の識別モデルを用いた埋め込みと後段の局所尤度を直列化し，異常データを効果的に活用できる設計を検討する．

3.3　提案手法

3.3.1　概要

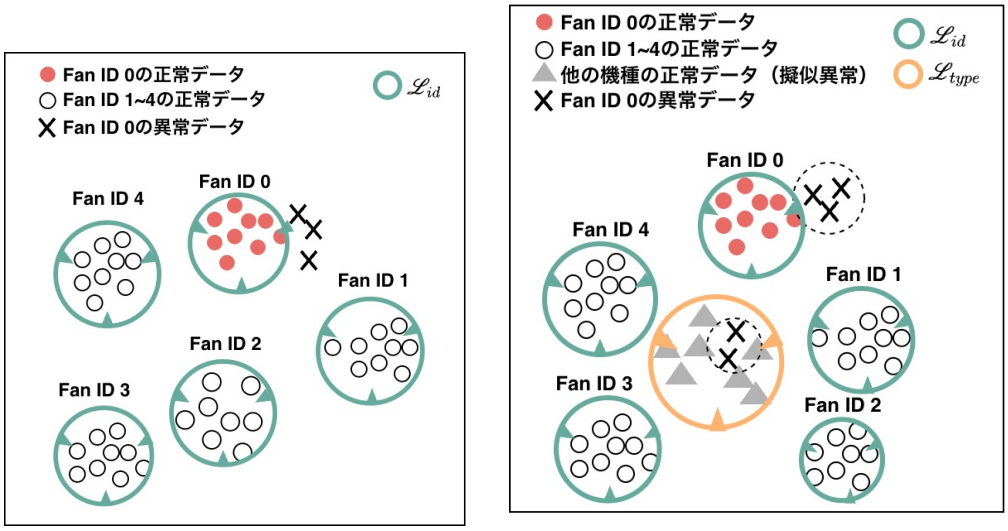
提案手法は異常データの収集は通常困難という背景から，異常音の収集が困難な場合は正常のみで学習し，利用可能な場合は少量の異常も取り入れて性能を高められる構成を目指した．そこで，畳み込みニューラルネットワークによる特徴量抽出器と GMM による異常検出器を直列に組み合わせる直列法に，OE による二値分類を組み込んだ Serial-Outlier Exposure（Serial-OE）を提案する．OE を取り入れることで，異常データの取り込みも容易になる．

図 3.1 に，従来の単一クラス学習と本手法における異常データ活用の概念的な差異を示す．図 3.1(a) に示すように，従来手法は正常データの分布のみをモデル化することに主眼を置いている．そのため，少量の実異常データが得られたとしても，それを「正常ではないもの」として学習プロセスに直接組み込むことは難しく，モデル構造の変更や複雑な正則化が必要となる場合が多い．

一方，図 3.1(b) に示す提案手法では，あらかじめ「正常」と「それ以外（疑似異常）」を識別する二値分類問題を解く枠組みを採用している．ここでは，損失関数 $\mathcal{L}_{\text{type}}$ により，正常データを原点から遠ざけ，異常（疑似異常）データを原点付近に引き寄せる

力が働く．この構造により，実際の異常データが得られた場合，それを新たなクラスとして定義する必要はなく，既存の「疑似異常」グループに追加するだけでよい．すなわち，擬似異常という受け皿があらかじめ用意されているため，実異常データをシームレスに統合し，正常データとの境界をより明確に洗練させることが可能となる．

図 3.2 に概略を示す．提案法では機械種ごとに 1 つの特徴量抽出器，機械の属性ラベル（ID）ごとに 1 つの異常検出器を学習する．DCASE2020 Task 2 [13] では 6 種の機械があるので特徴量抽出器は 6 個，ID は 41 個あるので GMM は 41 個学習する．特徴量抽出器の学習には OE を用いる．そのため，対象機械種の正常を正常データ，他機械種の正常を疑似異常と定義する．異常データが手に入った場合は疑似異常データと同様に扱う．



(a) 正常データのみを用いる従来手法　　(b) 提案手法（Serial-OE）

図 3.1: 従来手法と提案手法（Serial-OE）における埋め込み空間の比較概念図．(a) 正常データのみを用いる従来手法では，正常データの分布（$\mathcal{L}_{id}$ による分類）のみを学習するため，少量の異常データが得られても，それを損失関数に直接組み込む明確な場所がない．(b) 提案手法では，正常データと疑似異常データを識別する二値分類（$\mathcal{L}_{type}$）を導入している．これにより，異常データ（疑似異常および実異常）を原点付近に配置する受け皿が構造的に用意されており，実異常データが得られた際には，それを疑似異常クラスに追加するだけでシームレスに学習へ統合し，正常との境界を強化できる．

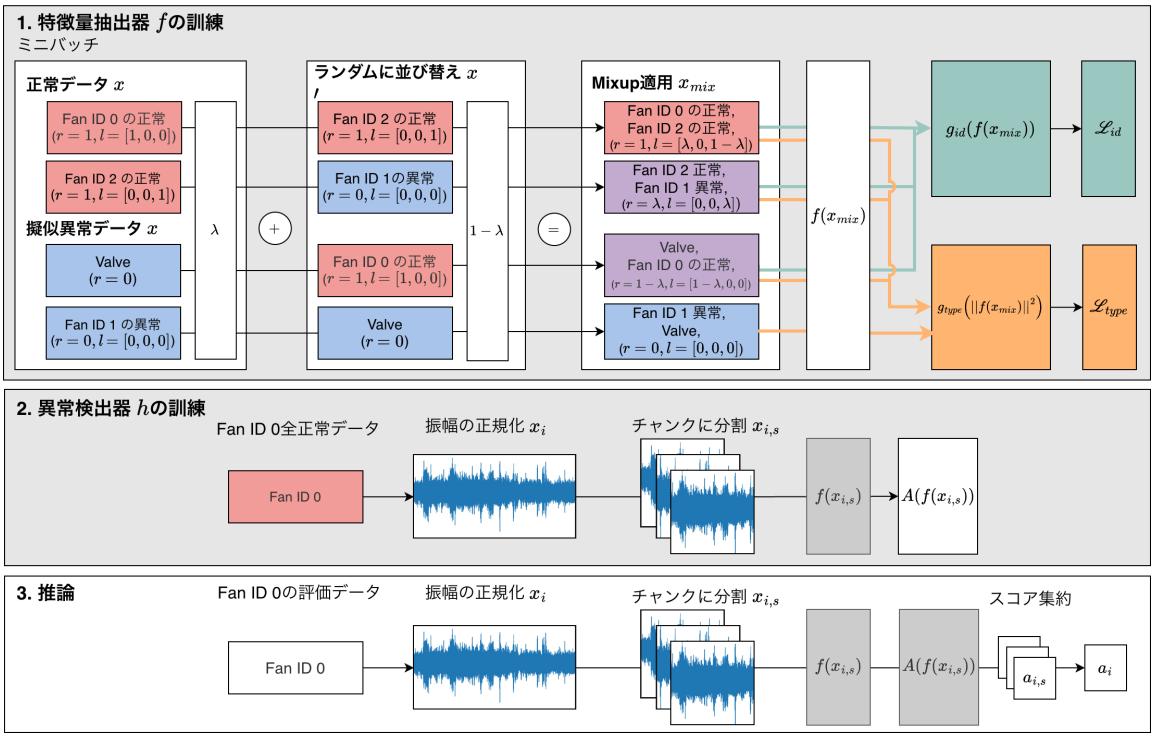


図 3.2: 提案手法 Serial-OE の概要．灰色の領域は学習過程，白色の領域は推論過程を表す．本例では，機械タイプ Fan の動作音に対して異常を検出する異常音検知システムの学習手順を示す．第一に特徴量抽出器 f を学習する．バッチサンプラによってミニバッチ内でサンプリングされた正常音と疑似異常音の混合音から抽出された埋め込みに対して，2 種類の損失関数を適用して学習する．第二に異常検出器 A を学習する．異常検出器 A は ID ごとに個別に学習するため，ここでは事前学習済みの特徴量抽出器 f から得られた Fan の ID 0 の正常音の埋め込みを用いて異常検出器を学習する．第三に訓練済みモデルを活用して異常スコアを得る．テストサンプルを複数チャンクに分割し，各チャンクを事前学習済みの特徴量抽出器 f と学習済みの異常検出器 A に入力して異常スコアを計算し，それらのスコアを集約して結果を得る．

異常音検知性能を下げる要因の一つは，環境音ノイズを異常 / 正常と誤判定することにある．よって環境音の影響を無視し動作音に着目できるよう学習させるのが望ま

しい．同一環境下で複数機械の動作音を識別するタスクで学習すると，モデルは環境音の影響を抑え，動作音差で分類するようになる [37]．擬似異常データとして他の機種の音を用いることで，対象機種の属性分類に関係のない音を無視するように学習するため，不要因子に対して頑健になり性能が向上する [37]．また，この選定に基づく正常と疑似異常の分類は性能を改善することが実験的に示されている [42], [46], [48]．なお，異常データが手に入った場合も異常検出器（GMM）の学習には正常データのみを用いる．以下に提案手法の詳細な定義を示す．

3.3.2　前処理

訓練データ N サンプル $\boldsymbol{x}_i \mid i = 1, \ldots, N$ を考える．各サンプル $\boldsymbol{x}_i$ には 2 種類のラベルが付与されているとする：

- 異常ラベル $y_i \in \{0, 1\}$：$y_i = 0$ は正常，$y_i = 1$ は疑似異常（対象機械種以外の正常音，あるいは対象機械種の実異常音を含む）
- 動作属性ラベル（ID） $l_i \in \{1, \ldots, C\}$

このとき，後続の記述を簡潔にするため，正常であることを表す指標

$$r_i := 1 - y_i \tag{3.1}$$

を定義する．すなわち $r_i = 1$ は正常，$r_i = 0$ は疑似異常を意味する．また，ID を one-hot 符号化した $\{0, 1\}^C$ 次元ベクトルを $\boldsymbol{l}_i$ とし，その c 番目の要素を $l_{i,c}$ と書く．

振幅は正常データ（$r_i = 1$）における平均と標準偏差を用いて標準化する．各 $\boldsymbol{x}_i$ が長さ L 秒の信号とすると，モデルに入力する際にはチャンク長を固定するために固定長の窓 S 秒で分割し，メルスペクトログラム（入力音響特徴） $\boldsymbol{x}_{i,s}$ を得て特徴量抽出器に入力する．

3.3.3 特徴量抽出器の損失関数

特徴量抽出器 f は，各チャンク $\boldsymbol{x}_{i,s}$ から D 次元埋め込み $\boldsymbol{z}$ を出力するネットワークである．提案手法では，この埋め込み空間が

1. 正常と疑似異常を大域的に分離できること
2. 正常データについては ID ごとの細かな差異を保持できること

を狙っており，その概念図を図 3.1(b) に示す．これを実現するため，二つの損失関数をマルチタスク学習で同時に最適化する．

一つ目の損失関数 $\mathcal{L}_{\text{type}}$ は，埋め込みのノルムを用いた「正常 vs 疑似異常」の二値分類 BCE である：

$$\begin{aligned}\mathcal{L}_{\text{type}} = -\frac{1}{N}\sum_{i=1}^{N}\Big\{ & r_i \log\big(g_{\text{type}}(\|f(\boldsymbol{x}_{i,s})\|^2)\big) \\ & + (1-r_i)\log\big(1-g_{\text{type}}(\|f(\boldsymbol{x}_{i,s})\|^2)\big)\Big\},\end{aligned} \tag{3.2}$$

ここで g_{type} はノルム $\|f(\boldsymbol{x}_{i,s})\|^2$ にアフィン変換とシグモイド関数を適用して，「正常である確率」を出力する関数である．g_{type} が大きいほど正常（$r_i = 1$）と判断されるよう学習される．このとき，正常は埋め込みのノルム $\|f(\boldsymbol{x}_{i,s})\|$ が大きい方向，擬似異常は小さい方向に押し出される．すなわち擬似異常は埋め込み空間の原点近傍の超球領域に集約され，正常は原点から離れた領域に分布するようになる．これは，擬似異常（他機種音・環境音・既知異常など）から得られる情報量を抑え，動作音に依存する本質的な特徴のみを残すことを意図している．さらに対象機械種の実異常を疑似異常側（$r_i = 0$）として加えることで，既知異常の特徴が原点近傍に集約され，後段の GMM による検知が容易になる．

二つ目の損失関数 $\mathcal{L}_{\text{id}}$ は，正常サンプル（$r_i = 1$）に限定して属性ラベル（ID）を識別する多ラベル BCE である．まず，各クラス c について，正常サンプルだけが正例

として寄与するターゲット

$$l'_{i,c} = r_i l_{i,c} \tag{3.3}$$

を定義する．これを用いると $\mathcal{L}_{\mathrm{id}}$ は

$$\begin{aligned}\mathcal{L}_{\mathrm{id}} = -\frac{1}{C\sum_{i=1}^{N} r_i}\sum_{i=1}^{N}\sum_{c=1}^{C}\Big\{ & l'_{i,c}\log\big(g_{\mathrm{id}}^{(c)}(f(\boldsymbol{x}_{i,s}))\big) \\ & + (1-l'_{i,c})\log\big(1-g_{\mathrm{id}}^{(c)}(f(\boldsymbol{x}_{i,s}))\big)\Big\},\end{aligned} \tag{3.4}$$

となる．ここで $g_{\mathrm{id}}^{(c)}$ は埋め込み $f(\boldsymbol{x}_{i,s})$ にアフィン変換とシグモイド関数を適用し，ID c に属する確率を出力する関数である．式 (3.3) により，$r_i = 0$（疑似異常）のサンプルは ID 学習の更新から除外され，$r_i = 1$（正常）のサンプルのみが各 ID の特徴をより厳密に表現するように働く．また ID を one-hot ではなくクラスごとに独立なシグモイドで推定することで，未知の ID が入力された場合にも「どの ID にも強く当てはまらない」という出力を許容でき，これにより誤検知の抑制が期待される．

最終的な損失関数は

$$\mathcal{L} = \mathcal{L}_{\mathrm{type}} + \alpha\mathcal{L}_{\mathrm{id}}, \tag{3.5}$$

であり，重み α はハイパーパラメータである．

3.3.4　ミニバッチサンプリング戦略

訓練を安定させるために，本研究ではミニバッチのサンプリング戦略を導入する．本研究で用いる正常 / 疑似異常の定義では，正常（$r_i = 1$）と比較して疑似異常（$r_i = 0$）は一般にサンプル数が多い．このままでは，ミニバッチ中で式 3.4 によって更新される正常サンプルが極端に少なくなり，$\mathcal{L}_{\mathrm{id}}$ の勾配が不安定になる．そこで，ミニバッチ内の正常：疑似異常の比率を常に 1:1 に保つバッチサンプラーを導入する．異常データが利用可能な場合は，ミニバッチ中の疑似異常の一部を実異常と差し替える（実異

常も $r_i = 0$ として扱う）．エポック更新は，正常サンプル（$r_i = 1$）を一巡したかどうかで定義する．これによりクラス不均衡を避けつつ学習時間も削減でき，$\mathcal{L}_{\mathrm{id}}$ の発散を防ぐ．

さらに，本研究では Mixup [68] を導入して決定境界をなめらかにし，連続的なゆらぎを含むサンプルに対してもロバストな識別を促す．Mixup は異常音検知でも広く使われている [37], [41]．ここで，サンプル $(\boldsymbol{x}_i, r_i, \boldsymbol{l}_i)$ と，同一バッチ内からシャッフルして選んだ $(\boldsymbol{x}_j, r_j, \boldsymbol{l}_j)$ を用い，$\lambda \sim \mathrm{Beta}(\beta, \beta)$ からサンプリングした $\lambda \in (0, 1)$ を用いて，Mixup 後のサンプルを

$$\boldsymbol{x}_i^{\mathrm{mix}} = \lambda \boldsymbol{x}_i + (1 - \lambda)\boldsymbol{x}_j, \tag{3.6}$$

$$r_i^{\mathrm{mix}} = \lambda r_i + (1 - \lambda) r_j, \tag{3.7}$$

$$\boldsymbol{l}_i^{\mathrm{mix}} = \lambda r_i \boldsymbol{l}_i + (1 - \lambda) r_j \boldsymbol{l}_j \tag{3.8}$$

と定義する．ここで r_i^{mix} は「この混合サンプルがどの程度正常であるか」を表す連続値（0〜1）であり，$\boldsymbol{l}_i^{\mathrm{mix}}$ は ID に対するソフトターゲットである．$\boldsymbol{l}_i^{\mathrm{mix}}$ は r_i と r_j によって重み付けしているため，疑似異常どうしの混合（$r_i = r_j = 0$）ではゼロベクトルになり，正常成分を含む混合では正常側 ID の割合を反映したソフトラベルになる．

この Mixup サンプルに対して，$\mathcal{L}_{\mathrm{type}}$（正常 vs 疑似異常）は

$$\mathcal{L}_{\mathrm{type}}^{\mathrm{mix}} = -\Big[r_i^{\mathrm{mix}} \log g_{\mathrm{type}}(\|f(\boldsymbol{x}_i^{\mathrm{mix}})\|^2) + (1 - r_i^{\mathrm{mix}}) \log\big(1 - g_{\mathrm{type}}(\|f(\boldsymbol{x}_i^{\mathrm{mix}})\|^2)\big) \Big] \tag{3.9}$$

と書ける．すなわち r_i^{mix} を「正常クラスのソフトターゲット」，$(1 - r_i^{\mathrm{mix}})$ を「疑似異常クラスのソフトターゲット」として扱う．

同様に，$\mathcal{L}_{\mathrm{id}}$（ID 判別）は，$\boldsymbol{l}_i^{\mathrm{mix}}$ をターゲットとする多ラベル BCE で

$$\mathcal{L}_{\mathrm{id}}^{\mathrm{mix}} = -\sum_{c=1}^{C} \Big[l_{i,c}^{\mathrm{mix}} \log g_{\mathrm{id}}^{(c)}(f(\boldsymbol{x}_i^{\mathrm{mix}})) + \big(1 - l_{i,c}^{\mathrm{mix}}\big) \log\big(1 - g_{\mathrm{id}}^{(c)}(f(\boldsymbol{x}_i^{\mathrm{mix}}))\big) \Big] \tag{3.10}$$

と表せる．ここで $l_{i,c}^{\mathrm{mix}}$ はベクトル $\boldsymbol{l}_i^{\mathrm{mix}}$ の c 番目の要素であり，$g_{\mathrm{id}}^{(c)}(\cdot)$ は ID c に属する確率である．$\boldsymbol{l}_i^{\mathrm{mix}} = \mathbf{0}$ 即ち正常成分を含まない混合の場合，この項の寄与は実質的に 0 となり，正常成分が含まれる場合（$r_i^{\mathrm{mix}} > 0$）のみ ID 情報による更新が行われる．

この設計をケース別に見ると直感的である：

- (1) 疑似異常どうし（$r_i = r_j = 0$）：$r_i^{\mathrm{mix}} = 0$ で「正常らしさ」は 0，$\boldsymbol{l}_i^{\mathrm{mix}} = \mathbf{0}$（ID 分類のターゲットなし）．このサンプルは $\mathcal{L}_{\mathrm{type}}$ のみで「疑似異常側」として学習される．
- (2) 正常どうし（$r_i = r_j = 1$）：$r_i^{\mathrm{mix}} = 1$，$\boldsymbol{l}_i^{\mathrm{mix}} = \lambda \boldsymbol{l}_i + (1-\lambda)\boldsymbol{l}_j$．正常サンプル同士の Mixup により，ID ごとの境界がなめらかに学習される．
- (3) 正常と疑似異常（例えば $r_i = 1, r_j = 0$）：$0 < r_i^{\mathrm{mix}} < 1$ となり「半分だけ正常」といった中間的な例を作れる．$\boldsymbol{l}_i^{\mathrm{mix}} = \lambda \boldsymbol{l}_i$ のように正常側の ID 情報だけが残る．

以上より，Mixup 後のサンプル $(\boldsymbol{x}_i^{\mathrm{mix}}, r_i^{\mathrm{mix}}, \boldsymbol{l}_i^{\mathrm{mix}})$ を用いることで，

- $\mathcal{L}_{\mathrm{type}}$ は常に更新され（正常 vs 疑似異常の連続的な境界を学べる），
- $\mathcal{L}_{\mathrm{id}}$ は $r_i^{\mathrm{mix}} > 0$（正常成分を含む）場合のみ寄与し，正常データ由来の ID 構造をよりスムーズに学習できる．

これにより，正常 / 疑似異常のクラス不均衡と，ID 付与の適用範囲の両方を制御しつつ，安定した学習が可能となる．

3.3.5　異常検出器の学習

学習済み特徴量抽出器から得られる D 次元埋め込みに対して，ID ごとに GMM を用いた異常検出器を学習する．具体的には，各 ID $c \in \{1, \ldots, C\}$ について，その ID に属する正常サンプル（すなわち $r_i = 1$ であるサンプル）のみを用いて GMM を異常検出器 A_c として推定する．ここで，1 つの長さ L 秒の録音を S 秒幅・50% オーバー

ラップで分割し，

$$M = \left\lceil \frac{2L}{S} \right\rceil$$

個のチャンク $\boldsymbol{x}_{i,m}$（$m = 1, \ldots, M$）を得る．各チャンクを特徴量抽出器 f に通し，

$$\boldsymbol{z}_{i,m} = f(\boldsymbol{x}_{i,m}) \in \mathbb{R}^D$$

を得る．ID c の GMM A_c は，その ID c に属する正常チャンク（$r_i = 1$ かつ $l_{i,c} = 1$）から得られる埋め込み $\boldsymbol{z}_{i,m}$ の集合に対して適用させる．ここでは，各録音から得られるチャンク数，および ID 全体で集約したサンプル数 M が次元 D に比べて十分大きいと仮定し，GMM の各混合成分には一般の共分散行列を用いる．

推論時には，対象機械種は既知であり，また DCASE2020 Task 2 の前提と同様に録音対象の ID も与えられるとする．テスト信号に対しても同様に S 秒幅・50% オーバーラップで M 個のチャンク $\boldsymbol{x}_{i,m}$ を切り出し，それぞれを f に通して埋め込み $\boldsymbol{z}_{i,m}$ を得る．各チャンク $\boldsymbol{z}_{i,m}$ に対して，対応する ID の GMM の対数尤度を計算し，その負値を異常スコアとする：

$$a_{i,m} = -\log p_{A_c}(\boldsymbol{z}_{i,m}), \tag{3.11}$$

ここで $p_{A_c}(\cdot)$ は ID c の正常分布を表す GMM の確率密度関数である．尤度が低い，すなわち $a_{i,m}$ が大きいチャンクは，その ID の正常パターンから外れているとみなされ，異常と判断されやすい．

3.3.6 異常スコアの集約

推論時には，各テスト音源を M 個のチャンクに分割し，それぞれからチャンク単位の異常スコア $a_{i,m}$（$m = 1, \ldots, M$）を得る．最終的に，音源全体として 1 つの異常スコア a_i を計算し，これを閾値判定に用いる．

異常の現れ方には大きく 2 種類がある．ひとつはモータの劣化など，ほぼ全区間にわたって持続的に異常が発生する「定常的な異常」，もうひとつは打撃音や擦過音のように，短い区間にのみ強い異常が出る「非定常的な異常」である．単純に全チャンクの平均をとると後者が埋もれやすく，逆に最大値のみをとると単発的なスパイク雑音に過敏になる．本研究では，2.4 節で述べた方針に基づき，両者に対応するための集約器を次のように定義する．

まず，チャンクスコア列

$$\{a_{i,1}, a_{i,2}, \ldots, a_{i,M}\}$$

を降順に並べ替えた列を

$$a'_i[1] \geq a'_i[2] \geq \cdots \geq a'_i[M]$$

とおく．次に，スコア列全体の中央値

$$\mathrm{med}_i = \mathrm{median}\big(\{a_{i,m}\}_{m=1}^{M}\big)$$

を計算する．上位半数に相当するインデックス数を

$$K = \left\lceil \frac{M}{2} \right\rceil$$

とすると，最終スコア a_i は

$$a_i = \frac{\sum_{m=1}^{K} \mathbb{1}[a'_i[m] > \mathrm{med}_i]\ a'_i[m]}{\sum_{m=1}^{K} \mathbb{1}[a'_i[m] > \mathrm{med}_i]}, \tag{3.12}$$

で定義する．ここで $\mathbb{1}[\cdot]$ は指示関数であり，条件が真のとき 1，偽のとき 0 を返す．すなわち，

- チャンクスコアが高いほうから上位 $K = \lceil M/2 \rceil$ 個だけを見る（異常が長時間続く場合には，その大部分がここに含まれる），

- その中でも全体の中央値 med_i より明らかに高いチャンクだけを平均する（短時間の強い異常ピークも拾える），

という二段階のふるい分けを行っている．なお，分母が 0，すなわち $a'_i[m] > \mathrm{med}_i$ を満たす上位チャンクが存在しない場合は，$a_i = \mathrm{med}_i$ とする実装とした．これは，全チャンクがほぼ同程度で顕著な異常が見られない場合，スコアが中央値に落ち着くようにするためである．

まとめると，式 3.12 により，

- 定常的な異常では，多くのチャンクで高いスコアが得られるため，上位半数の平均が高くなる，
- 非定常な突発異常でも，中央値より十分大きい一部のチャンクがあればそれらだけで平均が計算されるため，短い異常も見逃しにくい，

という性質を同時に満たすことができる．最終的な異常判定は，この a_i が閾値を超えるかどうかで決定される．実運用では，対象機械種および録音条件ごとに妥当な閾値を別途設定する必要がある．

3.4 実験評価

3.4.1 データセット

提案法の性能評価には，代表的な異常音検知データセットであり近年の異常音検知手法の評価にも用いられている DCASE2020 Task 2 データセットを用いた．類似のデータセットとして DCASE2021 Task 2，DCASE2022 Task 2，DCASE2023 Task 2, DCASE2024 Task 2, DCASE2025 Task 2 [9], [11], [12], [111], [112], [113], [114] があるが，これら後続のデータセットはドメインシフトの制御を目的としているため階層構造が必ずしも特定の対象機械の音声データ単位で分割されていない．この違いは，提

案法を評価し性能向上に寄与した要因を精査する際の議論を複雑にする．そこで本研究では，異常データの活用によって異常音検知性能を高めるという課題設定に焦点が合っている DCASE2020 Task 2 データを用いた．

DCASE2020 Task 2 は MIMII [115] と ToyADMOS [116] の 2 つのデータセットから構成され，いずれも機械動作音と環境音を含む．機械動作音は静穏環境下で複数の収録条件により収集され，環境音は実工場で複数条件のもと収集された．タスク主催者はこれら 2 つのデータセットを種々の SNR で混合して音声サンプルを生成している．各サンプルは長さ 10 秒，1 チャネル，16,000 Hz で収録されている．機械タイプは 6 種であり，MIMII から Fan，Pump，Slider，Valve，ToyADMOS から ToyCar と ToyConveyor を用いた．評価には，多くの先行研究 [1], [33], [36], [39], [40], [41] と同一の設定で DCASE2020 Task 2 の開発用データセットを用いた．モデルのハイパーパラメータは先行研究の値に基づいて選択した．機種タイプ内には属性を表す ID というラベルが含まれる．各 ID ごとに，学習用の正常音が約 1,000 サンプル，評価用として正常音と異常音がそれぞれ約 100 サンプルずつ含まれる．各機械タイプには 3〜4 種類の異常が存在するが，評価セットからは各異常音にどの種類の異常が含まれているかを知ることはできない．41 個の異常検出器 A を個別に学習する際には，各 ID に属するすべての正常データから得た埋め込みを，当該検出器の学習データとして用いた．

3.4.2　実験条件

議論を具体的にするため，本小節では機械タイプ Fan を例にモデル構築を説明する．なお，他の機械タイプについても同一のモデルを実装し，比較を簡素化するためすべて同一のハイパーパラメータを用いた．Fan の異常音検知モデルを構築する際，正常データには Fan の 7 つの ID すべての正常データを含めた．疑似異常データには，Pump，Slider，Valve，ToyCar，ToyConveyor のすべての正常音を用いた．

まず前処理として，Fan の正常データ全体の振幅の平均と分散を算出し，学習用データ

セット全体の振幅を正規化した．正規化後のデータを $S = 2.0$ 秒にランダム分割し，周波数ビン数 224，窓長 128 ms，ホップ 16 ms の Mel フィルタバンクを用いて 50–7,800 Hz の帯域で入力音響特徴を抽出した．学習用データのスペクトログラムを解析したところ，動作音のエネルギが高域に集中していることを確認したため，全機械でパワーが集中する 50〜7,800 Hz の周波数帯域に着目できるようバンドパス処理を施した．得られた入力音響特徴を，ImageNet [117] で事前学習された CNN ベースの特徴量抽出器 EfficientNet-b0 [118] に入力し，$D = 128$ 次元の埋め込みを得た．EfficientNet-b0 の学習では，学習率 0.001，最適化手法 AdamW [119]，スケジューラ OneCycleLR [120]，エポック数 100，ミニバッチサイズ 128，式 3.5 の $\alpha = 10.0$，Mixup のベータ分布 $\mathcal{B}(\beta, \beta)$ のハイパーパラメータは $\beta = 0.2$ とした．特徴量抽出器の学習後，異常検出器 A の GMM を学習した．GMM の混合数は 2 とした．推論時には，長さ $L = 10$ 秒の音源を $S = 2.0$ 秒の 10 セグメントに重複を許容して分割した．最後に，全セグメントに集約器を適用し，対象音源の異常スコアを算出した．

3.4.3 異常音検知性能の評価

提案法の性能を，提案法と性質の近い手法群および近年の最先端手法群と比較した．各手法の概要は以下のとおりである：

GMADE+SSC [15], [32] DCASE2020 Task 2 Challenge [13] の 1 位手法である．Mel スペクトログラムの各フレームに対して密度推定にマスク付きオートエンコーダを用いる手法（GMADE） [15] の異常スコアと，ID 分類とデータ拡張の 2 つの自己教師あり分類を行う手法（SSC） [32] の異常スコアをアンサンブルして最終スコアを算出する．

DDCSAD+BCE [42] 対象機械の対象 ID に属する正常データを「正常」として用い，それ以外の ID の正常データおよび他機械タイプの正常データを「疑似異常」と

して用いる手法である．なお,「正常データ」と「疑似異常データ」の定義は提案法とは異なる点に注意されたい．本手法は異常データを活用することで性能改善することができるためベースラインに加えた．

STgram-MFN [33] スペクトログラムと，1 次元 CNN で抽出した T-gram を用いた ID 分類から得られる特徴を利用する手法である．

SCAdaCos [1] 複数機械の ID に基づく分類で特徴量抽出器を学習し, 得られた埋め込みを GMM によって尤度を求める手法である．特徴量抽出器の学習には SCAdaCos を用い，Mixup と組み合わせて異常音検知に適した埋め込み空間を形成する．1 つのモデルで全機械の埋め込みを抽出し，異常スコア算出のために各 ID ごとに GMM を個別学習する．元論文に基づき 16 サブクラスタを用い，GMM の混合数も同数とした．その他の学習条件も原論文 [1] に準ずる．

SW-WaveNet [36] スペクトログラムに加えて WaveNet 由来のウェーブグラムを統合し，WaveNet 自体を特徴量抽出器として追加埋め込みを抽出する手法である．従来の生成モデルとしてではなく特徴量抽出器として WaveNet を用いる点が特徴であり，音声データからより詳細な特徴を抽出して解析能力を高める．

CLP-SCF [40] 二段階学習を採用する．第 1 段階ではコントラスト学習を用いて，機械タイプ間および同タイプ内の ID 間の関係性を活用して事前学習を行う．第 2 段階では自己教師あり分類を導入する CLP-SCF により微調整し，異常音検知に有用な音響特徴の学習を強化する．特徴量抽出器には STgram-MFN を用いる．

Noisy-ArcMix [41] Mixup で仮想合成したサンプルを ArcFace [2] により正常分布へ近づけて分類することで，クラス内分布のコンパクトさを高める．さらに，時間的アテンションブロックから導出した新しい入力特徴である temporal attention log-Mel spectrogram（TAgram）を STgram-MFN に導入する．

表 3.1: DCASE2020 Task 2 データセット使用時における比較手法の平均性能．数値は aAUC（AUC と pAUC ($p = 0.1$) の平均）を 5 回の乱数シードを用いて計算した際の平均と標準偏差を表す．

Methods	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Average
GMADE+SSC [15], [32]	80.65	83.27	93.41	91.21	92.72	73.29	86.27
DDCSAD+BCE [42]	83.42±0.26	83.01±0.57	87.27±0.97	97.90±0.83	87.72±0.42	59.72±0.97	83.17±1.58
STgram-MFN [33]	91.51	86.85	98.58	99.04	91.06	69.09	89.35
SCAdaCos [1]	86.63±0.34	90.96±0.46	99.08±0.17	93.33±1.08	95.44±0.06	70.93±0.32	89.39±0.23
SW-WaveNet [36]	94.54	84.98	96.77	98.14	92.85	74.70	90.33
CLP-SCF [40]	95.11	91.18	98.65	99.70	93.02	69.00	91.12
Noisy-ArcMix [41]	**96.83**	90.72	98.52	**99.85**	93.44	72.53	91.98
Serial-OE (Proposed)	93.29±0.07	**94.87±0.13**	**99.56±0.17**	99.19±0.10	**96.54±0.12**	**77.77±0.49**	**93.54±1.00**

各手法の平均的な性能の比較には，受信者動作特性曲線下面積（AUC）と pAUC（$p = 0.1$）の平均である平均 AUC（aAUC）を用いた．また，先行研究 [33], [40], [111] に倣い，ID 間の性能安定性の比較には，各機械タイプ内で最も AUC が低い ID の AUC を表す最小 AUC（mAUC）を用いた．表 3.1 と表 3.2 に各手法の性能を示す．GMADE+SSC [15], [32]，STgram-MFN [33]，SW-WaveNet [36]，CLP-SCF [40]，Noisy-ArcMix [41] の性能は各論文で報告された数値を用いた．DDCSAD+BCE [42]，SCAdaCos [1]，および提案法については，異なる乱数シードで 5 回実験し，その平均と標準偏差を示した．

表 3.1 より，提案法は全機械タイプで GMADE+SSC を上回った．複数手法を組み合わせたアンサンブル法よりも，単一システムで良好な性能が得られている．また提案法は，Valve と Fan を除くすべての機械タイプで，従来法を凌駕していることが分かる．Valve の動作音はクリック音であり，Fan の動作音はプロペラが空気を切る音であるため，両者の間にはほとんど共通性がない．最良の従来法である Noisy-ArcMix と比較すると，Valve の aAUC の差は 0.66 ポイントと小さく，aAUC 自体も 99.19% と比較的高い．Fan では aAUC の差が 3.54 ポイントと中程度であり，aAUC は 93.29% であった．Fan の音は風切り音であり，他機械に比べて動作音と環境雑音の識別が難しい．Mixup によって生成された音源と元の音源を明確に分離するような学習方法が

表 3.2: DCASE2020 Task 2 データセット使用時における比較手法の性能安定性．数値は mAUC（各機械タイプ内で最低性能の ID の AUC）を 5 回の乱数シードを用いて計算した際の平均と標準偏差を表す．

Methods	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Average
DDCSAD+BCE [42]	69.58±1.95	71.33±0.93	79.12±1.89	95.19±3.02	74.90±1.13	54.96±1.27	74.18±1.71
STgram-MFN [33]	81.39	83.48	98.22	98.83	83.07	64.16	84.86
SCAdaCos [1]	76.55±0.93	87.83±0.38	99.09±0.18	90.31±1.16	94.24±0.26	69.25±0.71	86.88±0.32
CLP-SCF [40]	88.27	87.27	98.28	99.58	86.87	65.46	87.62
Noisy-ArcMix [41]	**92.67**	91.17	97.96	**99.89**	88.81	68.18	89.78
Serial-OE (Proposed)	84.12±0.33	**94.17±0.24**	**99.54±0.17**	99.13±0.19	**94.81±0.42**	**70.48±1.41**	**90.38±1.39**

Fan のような環境雑音の識別が困難な機械タイプにおいて効果的な可能性はあるが，詳細な原因の解析は今後の課題である．一方で，Noisy-ArcMix と比較した平均 aAUC は 1.7% 改善している．提案法と性質の近い SCAdaCos と比較しても，全機械タイプで aAUC において提案法が上回った．Noisy-ArcMix，SCAdaCos，および提案法はいずれも，分類タスクに Mixup を用いて正常データの微細な変化を検出するよう学習する点は共通している．一方，提案法は「1 機械タイプにつき 1 モデル」を学習し，関連性の低い他機械タイプを疑似異常と見なすサンプリング戦略を採用する点が異なる．この戦略は，「全機械タイプを 1 モデルで学習する」戦略と比べて，監視対象機械に特有な背景雑音を無視するだけでなく，監視対象機械に特化することで混合音を明示的に異常として識別できるという点で有効であると考えられる．

表 3.2 は，ID 間にわたる各手法の性能安定性を示す．たとえ aAUC が高くても，mAUC が低ければ特定の ID において一部の異常データを検出できていないことを意味する．したがって，mAUC が高い手法は性能が安定していると評価できる．安定性評価においても最良の従来法である Noisy-ArcMix と比べ，提案法は平均 mAUC で 0.7% 高い値を達成した．また，提案法と近縁の SCAdaCos と比較しても，全機械タイプで mAUC においてより高い安定性を示した．

以上の結果より，提案法は平均性能のみならず，多くの ID にわたって良好な性能を

示すことが確認できる．表 3.1 および表 3.2 に示した手法のうち，DDCSAD+BCE と Serial-OE を除く手法は，正常データのみが利用可能であることを前提としている．したがって，仮に異常データが取得可能であっても，それらを活用するには追加の工夫が必要となる．これに対し提案法は，異常データを容易かつ効率的に取り扱える点で優位である．この点については，3.4.6 節で詳述する．

3.4.4 アブレーションスタディ

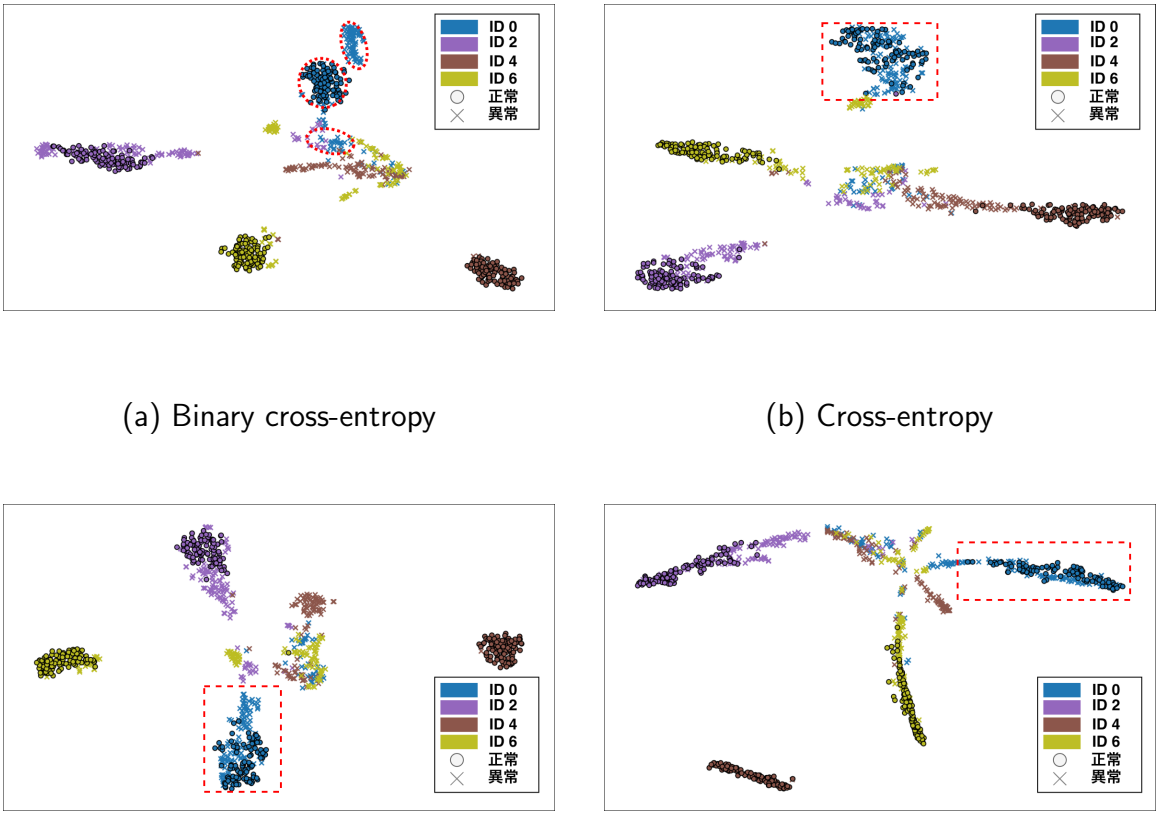
提案手法が高い性能を達成した要因を明らかにするため，各異常音検知手法の平均性能を比較する評価指標として aAUC を用い，アブレーションスタディを実施した．定性的評価のため，埋め込み空間を t-SNE [121] によって可視化した．

ID 分類に用いる損失関数の評価

表 3.3 は，各損失関数で特徴量抽出器を学習したときの aAUC を示す．図 3.3 は，Fan の DCASE2020 Task 2 開発セットにおいて，各損失関数を用いた場合に得られた埋め込み空間の t-SNE 可視化（perplexity= 40）を示す．表 3.3 および図 3.3 は，式 3.4 における $\mathcal{L}_{\mathrm{id}}$ に最も適した損失関数を比較したものである．GMADE+SSC はアンサンブル手法であるため，解析から除外した．提案手法および DDCSAD+BCE は ID 分類の損失関数として BCE を用い，SCAdaCos は ID 分類の損失関数として SCAdaCos を用いる．その他の手法（STgram-MFN，SW-WaveNet，CLP-SCF，Noisy-ArcMix）は，ID 分類の損失関数として ArcFace を用いる．

表 3.3: DCASE2020 Task 2 データセットを用いた，提案手法の $\mathcal{L}_{\mathrm{id}}$ 以外の要素に関するアブレーションの平均性能．値は aAUC（AUC と pAUC ($p = 0.1$) の平均）を 5 回の乱数シードを用いて計算した際の平均と標準偏差を表す．

Loss function	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Average
BCE (Proposed)	**93.29±0.07**	**94.87±0.13**	**99.56±0.17**	**99.19±0.10**	**96.54±0.12**	**77.77±0.49**	**93.54±1.00**
Cross-entropy	88.72±0.21	91.42±0.38	98.60±0.25	96.37±0.66	92.12±0.79	69.29±0.32	89.42±0.18
SCAdaCos [1]	84.83±0.52	88.72±0.59	97.81±0.68	88.97±2.68	84.11±0.79	63.38±1.01	84.64±0.46
ArcFace [2]	85.09±0.19	85.00±0.83	91.18±4.77	91.43±2.58	66.33±2.12	65.50±0.60	80.75±1.04



(a) Binary cross-entropy　(b) Cross-entropy

(c) SCAdaCos　(d) ArcFace

図 3.3: Fan の埋め込み空間の t-SNE 可視化（$\mathcal{L}_{\mathrm{id}}$ を変更）．(a) BCE, (b) Cross-entropy, (c) SCAdaCos [1], (d) ArcFace [2]．○ は正常，× は異常．ID 0 の正常・異常分布を赤破線で囲む．

本研究では, 式 3.4 に従って, 従来手法で用いられてきたこれらの ID 分類用損失関数を提案手法に適用して比較した．具体的には，式 3.4 の $\mathcal{L}_{\mathrm{id}}$ において用いている BCE を Cross-entropy，SCAdaCos，ArcFace に置き換え，得られる埋め込み空間の差異を評価した．なお，すべてのバージョンで Mixup と式 3.2 の $\mathcal{L}_{\mathrm{type}}$ も適用した．式 3.4 の損失関数を変更したことの影響を受けるサンプルは，(i) 正常データと正常データの混合，(ii) 正常データと疑似異常データの混合，の 2 種類の音を含む．以下では，これらのサンプルに着目し，各損失関数の効果を議論する．

BCE は，入力音に対象 ID の音が含まれているか否かに着目して特徴を獲得する．一方，Cross-entropy，SCAdaCos，ArcFace は，入力サンプルが属するクラスに基づいて分類するための特徴を獲得する．さらに ArcFace と SCAdaCos はクラス内分散の最小化を図り，SCAdaCos はサブクラスタを用いることで，複数の異なる特徴を持つ ID を各特徴に応じて分類できる．これら 3 つの損失関数では，損失関数の性質上，各クラスの確率の総和が 1 になることを要求する．そのため，正常データと疑似異常データを混合した場合，正常データの確率を 1 とした．正常データ同士の混合作成でも，各サンプルの通常の混合比を用いた．

表 3.3 の結果から，提案手法における $\mathcal{L}_{\mathrm{id}}$ の損失関数として BCE を用いたとき，すべての機械タイプで最良の異常音検知性能が得られたことがわかる．まず，BCE が Cross-entropy より高性能であった理由を議論する．図 3.3a（BCE）と図 3.3b（Cross-entropy）の可視化を比較すると，赤の点線で囲った ID 0 において，BCE の方が正常データのクラスがよりタイトにクラスタリングされていることがわかる．損失関数の性質を考えると，Cross-entropy のように ID を分類するための特徴よりも，特定の ID が含まれているかどうかに着目して得られた特徴の方が，異常音検知 に有効であると考えられる．

次に，$\mathcal{L}_{\mathrm{id}}$ の損失関数として SCAdaCos と ArcFace を用いた場合の性能を比較する．表 3.3 に示すとおり，SCAdaCos または ArcFace を用いると異常音検知性能は低

下した．図 3.3a（BCE）と，図 3.3c（SCAdaCos），図 3.3d（ArcFace）を比較すると，SCAdaCos と ArcFace の両方で，ID 0（赤の点線）における正常データのクラスタ内に異常データが多く含まれていることがわかる．SCAdaCos と ArcFace は，埋め込みのノルムを用いて $\mathcal{L}_{\text{type}}$ による分類も行うため，角度ベースの分類を行う難易度が増したことが，性能低下の一因である可能性がある．そのため，埋め込みのノルムを用いる損失関数と角度ベースのメトリック学習の組み合わせには注意が必要である [37]．

以上の結果から，特徴量抽出器を学習する際にどの損失関数を用いるかは，入力音に対象機械の音が含まれているかどうかを識別するうえで重要であることが示唆される．

特徴量抽出器の学習手法の評価

表 3.4 は，特徴量抽出器の学習時に導入した各要素を順に削除した場合の提案手法の性能比較を示す．まず平均に着目すると，いずれかの要素を削除すると提案手法に比べて性能が低下し，かつ，$\mathcal{L}_{\text{id}}$ のみの場合と比べてすべての手法で性能が向上していることから，これら 3 要素はいずれも有効であることが示唆される．さらに，2 要素を組み合わせた場合の性能と各単独要素の性能を比較すると，すべての組み合わせで同等以上の性能向上が得られた．各要因の異常音検知性能への影響を比較した結果，性能改善への寄与が大きい順に，Mixup，ImageNet による重みの事前学習，損失関数 $\mathcal{L}_{\text{type}}$ であった．

Mixup によって異常音検知性能が大幅に改善したことは，埋め込み空間の中間表現を獲得することが性能向上に有効であることを示している．これにより，正常データからわずかに異なる特徴を持つ異常データの分布を，正常データからより離れたクラスタへと配置できる．

また，ImageNet の分類タスクで事前学習されたモデルの重みで特徴量抽出器を初期化することにより，すべての機械タイプで異常音検知性能が改善した．この知見は，事前学習モデルの利用による性能向上を報告している先行研究 [110], [122] と整合的であ

表 3.4: DCASE2020 Task 2 データセットを用いた，提案手法の $\mathcal{L}_{\mathrm{id}}$ 以外の要素に関するアブレーションの平均性能．値は aAUC（AUC と pAUC ($p = 0.1$) の平均）を 5 回の乱数シードを用いて計算した際の平均と標準偏差を表す．

Method	L_{type}	weight	Mixup	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Average
Serial-OE	✓	✓	✓	**93.29±0.07**	**94.87±0.13**	**99.56±0.17**	**99.19±0.10**	**96.54±0.12**	**77.77±0.49**	**93.54±1.00**
w/o L_{type}		✓	✓	90.46±0.27	92.69±0.36	98.66±0.29	98.01±0.45	94.59±0.32	74.44±1.67	91.48±0.35
w/o weight	✓		✓	87.15±0.33	87.96±0.57	98.70±0.19	95.19±0.68	91.95±0.56	72.50±0.22	88.91±0.09
w/o Mixup	✓	✓		86.52±0.37	91.98±0.76	98.40±0.32	98.83±0.58	94.30±0.44	62.40±0.74	88.74±0.22
w/ L_{type}	✓			85.15±0.31	85.86±0.76	98.76±0.24	96.41±0.43	91.96±0.22	58.28±0.28	86.07±0.11
w/ weight		✓		83.08±0.55	84.93±0.20	97.08±1.01	99.02±0.39	92.07±0.70	61.16±0.58	86.22±0.35
w/ Mixup			✓	87.05±0.66	88.12±0.40	97.48±0.28	98.91±0.13	91.48±0.63	70.76±0.31	88.97±0.12
only L_{id}				82.02±0.54	84.80±0.34	97.77±0.61	99.34±0.14	92.44±0.29	56.80±0.63	85.53±0.11

る．この結果は，他の分類タスクで学習されたモデルの重みを初期値として効果的に活用できることも示している．

最後に，損失関数 $\mathcal{L}_{\mathrm{type}}$ の効果に着目すると，$\mathcal{L}_{\mathrm{type}}$ をノルムに基づく正規・疑似異常の分類に用いることで 異常音検知性能が向上した．これは，$\mathcal{L}_{\mathrm{type}}$ により異常データが原点近傍に集中し，異常検知性能が改善するためである．さらに，$\mathcal{L}_{\mathrm{type}}$ は Mixup と併用することで，正常と疑似異常の中間表現も形成でき，これも性能向上に寄与すると考えられる．

これらのアブレーションの結果から，提案手法における Mixup，事前学習重みによる初期化，そして $\mathcal{L}_{\mathrm{type}}$ は，異常音検知 性能を向上させるうえでいずれも有効であることが示された．

ID 情報を用いない場合の評価

従来法および提案法は，学習時に ID 情報を用いることを前提としている．これは，対象機械の正常データをより詳細に反映した埋め込みを抽出でき，異常音検知 性能が向上するためである．したがって，ID あるいは同等の情報が利用可能であれば望まし

表 3.5: DCASE2020 Task 2 データセットを用いた，ID 情報なしで学習したモデルの平均性能．値は aAUC（AUC と pAUC ($p = 0.1$) の平均）を 5 回の乱数シードを用いて計算した際の平均と標準偏差を表す．

Method	$\mathcal{L}_{\text{id}}$	$\mathcal{L}_{\text{type}}$	Mixup	Fan	Pump	Slider	Valve	ToyCar	ToyConveyor	Average
Serial-OE	✓	✓	✓	**93.29±0.07**	**94.87±0.13**	**99.56±0.17**	**99.19±0.10**	**96.54±0.12**	**77.77±0.49**	**93.54±1.00**
SCAdaCos [1]	—	—	✓	86.63±0.34	90.96±0.46	99.08±0.17	93.33±1.08	95.44±0.06	70.93±0.32	89.39±0.23
SCAdaCos (w/o ID)	—	—	✓	75.72±0.51	77.10±0.61	87.44±0.43	76.25±0.55	76.79±0.82	54.83±0.27	74.69±0.38
w/o $\mathcal{L}_{\text{id}}$		✓	✓	77.86±0.94	81.93±1.61	88.54±1.65	71.77±1.21	69.66±1.79	54.52±0.85	74.05±0.45
only $\mathcal{L}_{\text{type}}$		✓		75.20±1.28	71.62±1.62	91.81±0.43	58.84±1.19	61.11±1.86	52.91±0.51	68.58±0.50

いが，常に利用できるとは限らない．表 3.5 は，特徴量抽出器と異常検出器のパラメータを ID 情報なしに学習した場合の異常音検知性能を示す．提案手法では，式 3.5 において $\lambda = 0$ とした．SCAdaCos では ID 情報を用いると 41 クラスの分類となるが，機械タイプのみを用いる場合は 6 クラスとなる．各手法において，機械タイプごとに GMM を学習して 6 つの異常検出器を作成した．表 3.5 の結果から，ID 情報を用いない場合，両モデルともすべての機械タイプで異常音検知性能が低下することがわかる．

ID を用いた場合と同様に，Mixup により正常と疑似異常の中間表現を獲得するように特徴量抽出器を学習すると異常音検知性能が向上し，中間表現の獲得が有効であることが確認できた．

次に，ID を用いない提案手法と，ID を用いない SCAdaCos の性能を比較したところ，平均性能の観点では両者はほぼ同等であった．正常データのみを用い，ID が利用できない場合には，両手法に差はない．しかし，提案手法は次節で述べるように，異常データを学習に活用できる利点がある．

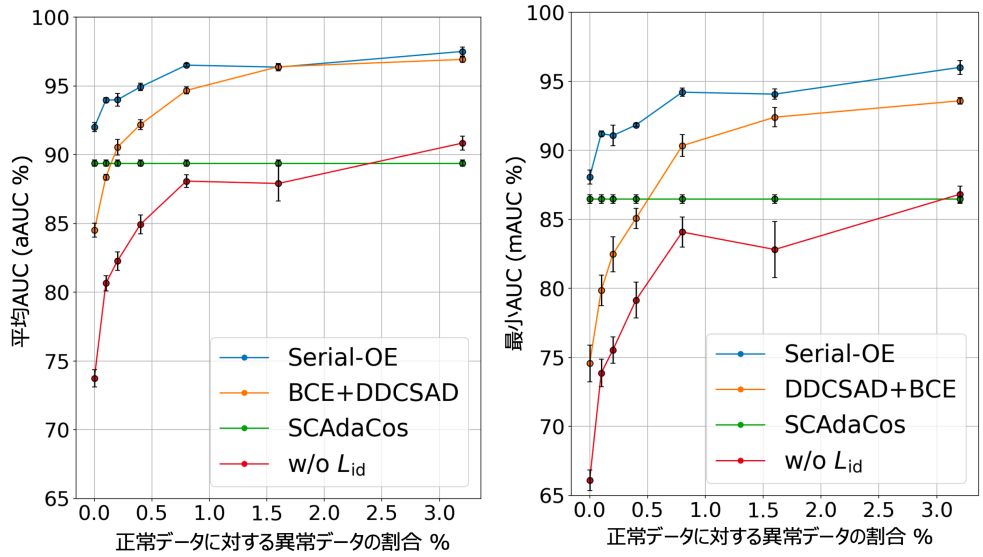
3.4.5 学習時に異常データを疑似異常データとして利用した場合の性能評価

実際の異常音データは，異常状態の検出に有用な特徴を含むため，性能向上に活用できる可能性が高い．そこで，少量の異常データを学習に用いることで，提案手法がより高い異常音検知性能を達成できるかを検討した．また，ID などの情報が利用できない場合における異常データ利用の有効性も評価した．

図 3.4a は，横軸に学習データに含める異常データの割合，縦軸に平均 AUC（aAUC）をとったグラフである．図 3.4b は同様に，縦軸を最小 AUC（mAUC）としたグラフである．ここで，ID あたりの学習データ量は 1,000 サンプル（各 10 秒）であるため，異常データ 1% は 100 秒に相当する 10 サンプルを意味する．実験では，異常データのサンプル数として $\{2^i \mid i = 0, 1, \ldots, 5\}$ を用いた．評価データには約 100 サンプルの異常データが含まれるため，十分な評価データ量を確保する目的で $i = 5$ を上限とした．提案手法は，節 3.3 で述べたように正常・疑似異常の二値分類を定義しているため，少量の実異常データであっても，利用可能なら性能向上に寄与する．

同じ実験を ID 情報が利用できない場合にも実施した．比較対象として，正常・疑似異常の二値分類を行う従来法である DDCSAD+BCE [42] を用いた．参考として SCAdaCos [1] の性能も示す．図 3.4a および図 3.4b から，aAUC と mAUC のいずれにおいても，Serial-OE が最高性能を達成し，次いで DDCSAD+BCE が高いことがわかる．これらの結果は, 平均性能と性能の安定性の両面で, 提案手法が DDCSAD+BCE よりも効果的に異常データを活用できることを示している．

提案手法は，異常データの 0.1%（10 秒の 1 サンプル）を学習用の正常データに追加しただけでも異常音検知性能を有意に改善し，少量の異常データを学習に用いる有用性を示した．この改善は aAUC の観点で 2.4% の向上に相当し，新たな異常音検知手法を開発するよりも容易に性能を引き上げる手段である．



(a) 平均 AUC（aAUC）　　(b) 最小 AUC（mAUC）

図 3.4: 学習に用いる異常データの割合と異常音検知性能の関係．（a）aAUC [%]，（b）mAUC [%] で評価．エラーバーは，異なる乱数シード 5 回の計算から得た標準偏差を表す．

次に，学習に異常データを用いるがIDが含まれない場合の異常音検知性能を比較した．図 3.4a と図 3.4b の結果から，異常データの割合が 3.2%（すなわち 320 秒の異常データ）であれば，提案手法の性能はIDを用いる従来法（SCAdaCos）と同等以上となることがわかる．多くの異常音検知手法は，異常音検知 に用いる音データの基礎的属性としてID情報が利用できることを前提としている．しかし実環境では，IDに相当するデータを常に収集できるとは限らない．一方で，監視対象の機械はある程度の頻度で異常状態を呈するため，異常データは入手可能であることが多い．さらに，学習には長時間の異常データは不要であり，数秒から数分程度でも性能向上に十分である．したがって，ID情報が利用できない場合でも，少量の異常データを用いることで性能を改善できる点は，提案手法の大きな運用上の利点である．加えて，各機械ごとのハイパーパラメータ調整を要さないことも，例えば多数の機械を監視する場合に容

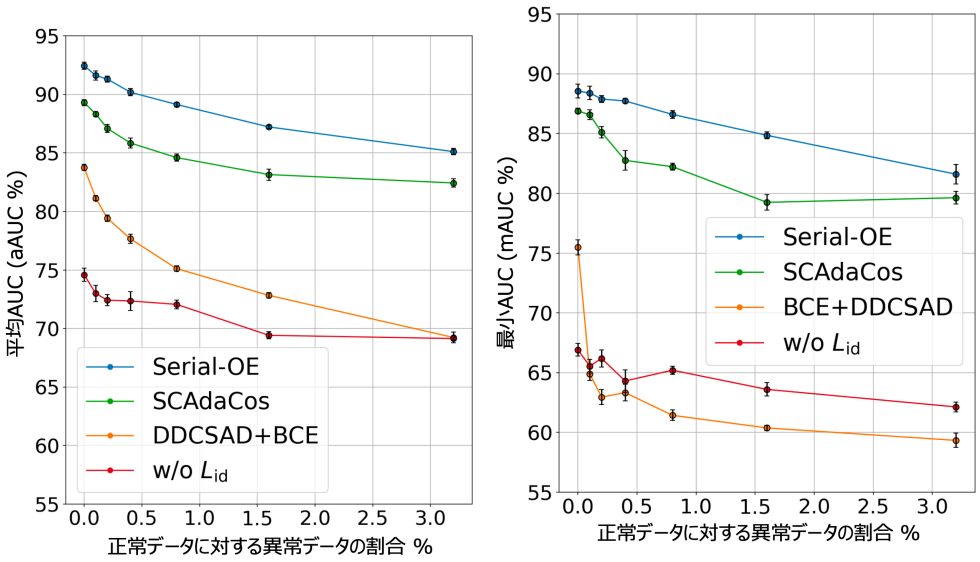
易にモデルを適用できるという運用上の利点である．

3.4.6 学習に用いる正常データに異常データが混入している場合の性能評価

実運用で収集した少量の異常データを異常音検知システムの学習に活用する際，収集方法によっては，その異常データが正常データを汚染してしまう場合がある．そこで，正常データに少量の実異常データが混入した場合の提案手法の性能を評価した．正常学習データに異常データをランダムに混入させ，そのデータで特徴量抽出器と異常検出器を学習した．図 3.5a は，横軸に混入した異常データの割合，縦軸に aAUC をとったグラフである．図 3.5b も同様に，縦軸を mAUC としたグラフである．図 3.5 から，異常データの混入割合が増えるほど，すべての異常音検知手法で性能が低下することがわかる．一方で，提案手法は他手法と比較して全混入水準で最良性能を達成し，他手法よりも汚染に対して頑健であることが示された．特に，正常データが約 80 秒分の異常データで汚染されている場合でも，提案手法は SCAdaCos と同等の性能を達成できた．もしこの程度の汚染異常データを適切に分類して学習に利用できれば，図 3.4a に示すように aAUC ベースで約 5% の性能改善が見込める．したがって，異常データの管理はモデル性能の向上という観点から非常に重要である．

異常データの混入割合に対する性能低下の相対比で見ると，DDCSAD+BCE の性能低下が最も急峻であった．DDCSAD+BCE は 1 つの ID につき 1 つのモデルを学習するため，汚染データの影響を受けやすいと考えられる．提案手法と SCAdaCos は，他手法よりも異常データの混入に対して感度が低かった．提案手法と SCAdaCos を比較すると，SCAdaCos に比べて，提案手法は混入割合が増加しても性能低下がより緩やかであった．提案手法は 1 つのモデルで 1 機械タイプの推論を行うのに対し，SCAdaCos は 1 つのモデルで全機械タイプの推論を行う．複数の機械タイプに対して埋め込み空

間を定義することで，各IDにおける正常データ汚染の影響が低減される可能性がある．この結果は，複数機械タイプに対して単一モデルを用いる方が，機械タイプごとに別モデルを用いる手法よりも異常データ汚染に対して頑健である場合があることを示唆している．一方で，異常データを活用した際に最も性能を向上させるのは機械タイプごとに別モデルを用いる手法であり，汚染耐性と性能向上の関係にはトレードオフが存在すると考えられる．



(a) 平均 AUC（aAUC）　　(b) 最小 AUC（mAUC）

図 3.5: 学習に用いる異常データが様々な割合で正常データに混入している場合における，異常データ割合と異常音検知性能の関係．（a）aAUC [%]，（b）mAUC [%] で評価．エラーバーは，異なる乱数シード 5 回の計算から得た標準偏差を表す．

3.5　制約事項

本研究では，実際の異常データを学習に用いた場合の異常音検知性能の変化を評価したが，現実には多様な種類の異常状態が存在する．DCASE2020 Task 2 データセッ

トの制約により，提案手法の学習において，どのような種類の異常データが性能向上に有効であるかは十分に検討できていない．提案する Serial-OE を導入する際の最適条件を見つけるためには，この点に関するさらなる研究は有用である．

さらに，ドメインシフトが性能に与える影響や，提案手法のエッジデバイスへの適用可能性についても，本研究の範囲外である．これらを検討することは有益であると考える．

3.6 結論

異常音検知 性能向上に関する従来研究の多くは，正常データの活用に主眼を置いてきた．本研究では，監視対象の機械がある程度の頻度で異常状態を示すという前提に基づき，学習に用いる少量の異常データの活用を検討した．本研究では，Outlier Exposure と直列法を用いる Serial-OE という新しい異常音検知手法を提案した．提案手法は，正常・疑似異常データの利用時，さらに少量の実異常データを疑似異常データと組み合わせて利用した場合のいずれにおいても異常音検知性能を改善し，評価実験で最高性能を達成した．

本研究ではまた，学習時の損失関数，可視化，ID 情報の影響，正常データが異常データで汚染された場合の性能など，提案手法の各構成要素に関する性能評価も行い，異常音検知 性能の改善や様々な条件下での提案手法の適用に関する有益な示唆を得た．

本研究は，DCASE2020 Task 2 データセットの正常データを用いた従来法との性能比較にとどまらず，実運用を想定した異常データの実用的な活用による異常音検知性能のさらなる向上にも焦点を当てた．異常データを利用すれば検出性能が向上すること自体は自明に見えるかもしれないが，機械の故障音 10 秒というごく少量のデータであっても有意な性能向上が得られることは，自明ではない．本研究は，異常データの収集は現実的でないという長年の見解を認識しつつも，そのようなデータを活用する

ことで得られる利点を示した.

第4章
未ラベル条件下における疑似異常データ集合の選択と疑似ラベル活用による異常音検知の改善

4.1　はじめに

本章は，異常ラベル y を用いず，機器の微細な状態や個体差に対応する属性ラベル（例：ID，運転設定）も用いない前提での異常音検知を扱う．この課題は表 1.1 における C の領域を対象とする．これは監視対象の機械の動作音のアノテーションがコスト面もしくは物理的制約など理由によって困難な状況を想定する．一方で，機械タイプという粗い区別は実運用でも容易に取得できるため，学習時の補助情報として許容する．この制約下で動作するモデルを開発することによって異常音検知システムの導入に対する制約を軽減することが期待される．

第 1 章，第 2 章で整理したとおり，本研究は前段で識別モデルを用いて不要因子（背景雑音 n・伝達系 h）に頑健な埋め込みを獲得し，後段で局所距離や密度等に基づく異常スコアを算出する直列法を基本とする．本章では，機械タイプのみ利用という最小の属性ラベルのもとで直列法を成立させる方法を体系化し，既存研究の限界を明確化した上で，機械タイプのみを用いつつ欠落している属性の効果を近似・補完するため

の提案手法を示す [123], [124]．本章の前提は，学習時と運用時の条件が異なるドメインシフト環境でも，明示的な属性ラベルなしで検知性能を維持したいという要求である．そこで本章では，(i) ソースおよびターゲットドメイン別クラスタ＋類似度ベース推論，(ii) 外部データからの擬似異常サンプル選別，(iii) 擬似ラベルの反復洗練という3ステップで，ドメインシフト下でも識別的な埋め込みを維持・改善する戦略を示す．詳細な数値結果は付録にまとめる．

4.2　関連研究：属性ラベルを用いない異常音検知

本章では，異常ラベルや属性ラベルを用いず，機械タイプのみが利用可能な設定特有の課題と，それに対して直列法がどのように拡張されてきたかを述べる．

ラベルなし設定では，機械タイプ内で運転条件や個体差に起因するサブクラス構造が混在しやすく，単一中心への収束や局所マージン不足によって，後段の局所距離スコアが鈍ることが報告されている [3]．この問題に対して識別前段を拡張する既往研究は大きく二つの方向に集約できる．

第一は，機械タイプのみの弱い監督のもとで角度幾何を整形しつつ，多峰性を過度に潰さない損失設計である．ArcFace / AdaCos / SCAdaCos などのマージン系は異常音検知で有効だが，ラベルが限定的な条件ではクラス内の多峰性を維持する設定が重要になる [37], [125]．Fujimura らは多解像度入力に対し，連結埋め込みと各ブランチ埋め込みの双方へ SCAdaCos を課す総和を サブスペース損失として実装し，機械タイプのみの設定でも識別モデルによって得られる特徴量を安定化できることを示した [3]．

第二は，属性の欠落で失われるサブクラス構造を，擬似ラベルで近似する方向である．具体的には，埋め込み空間のクラスタリング等で擬似ラベルを作り直し，機械タイプ内の潜在モードを顕在化させる [3]．ただし擬似ラベルには誤りが含まれるため，単にクラス内分散を縮めるだけでは環境ノイズなど誤ったクラスタに基づいて学習が進

む可能性がある．このため，ラベルなしでも小規模で適用できる距離学習を併用し，運転音の微少差を強調しつつノイズ差分を抑える学習圧を与える設計が提案されている．

ラベルなし下で外部データを利用する試みとして，擬似異常を外部から注入する OE がある [46], [90]．しかし無選別に話声や楽器音などを混在させると，機械音と関係のない境界を学習して劣化し得る．近年は埋め込み空間で異常をシミュレーションするアプローチも提案されるが [126]，直列法との整合という観点では，機械タイプ別に正常近傍として有用な外部データだけを選別し，投入量を制御する設計が望ましい．

正規化フローや再構成オートエンコーダによって属性ラベルなしの設定で異常音検知する方法も提案されているが，未観測の環境や伝達特性の揺らぎに対して尤度や再構成誤差が体系的にシフトしやすく，単一閾値運用に難しさがある [127]．ラベルなし設定では閾値の自動設計やスコア整合が特に重要になるため，本研究の評価条件では直列法の方が運用容易性と頑健性のバランスが取れると判断する．

以上を踏まえ，本研究ではベースラインとして，多解像度入力に対し連結および各ブランチ埋め込みへ SCAdaCos を課すサブスペース損失に基づく前段と，kNN を基本とする後段を採用する [3]．そのうえで，ラベルなし特有の欠落を補うため，次の三点を提案する．第一に，基準モデルの機械タイプ別最大正常スコアを閾値として用い，正常に近い外部データのみを擬似異常集合として選別し，外部データの無選別注入を避けるために投入量を上限で制御する．第二に，擬似ラベル付与に距離学習を併用して誤擬似ラベルの影響を抑え，運転音の微少差を強調する．第三に，これらを反復学習で洗練し，選別精度と擬似ラベルの質を段階的に高める．この設計により，属性ラベルなしでもサブクラス構造の回復と外部データの有益注入を両立させることを目指す．

4.3　ベースライン：機械タイプのみを用いた多解像度直列法

本節では，属性ラベルなしの異常音検知でもっとも高い性能を示している Fujimura らの多解像度ベースライン [3]（以下，Ba）を中心に，その骨格となる Wilkinghoff らの枠組み [37], [125] を取り入れた標準的設定を整理する．

4.3.1　ネットワークと入力音響特徴

各音声クリップは，互いに補完的な三つの入力音響特徴に変換する．(i) 窓長の長い STFT による 2 次元振幅スペクトログラム，(ii) 窓長の短い STFT による 2 次元振幅スペクトログラム，(iii) 全区間 DFT による 1 次元振幅スペクトルである．各入力音響特徴は専用の CNN ブランチに入力され，それぞれ 128 次元の埋め込みを出力する：

$$\boldsymbol{z}^{(m)} \in \mathbb{R}^{128}, \quad m = 1, 2, 3.$$

三つの埋め込みを連結して 384 次元の連結埋め込みを得る：

$$\boldsymbol{z}^{\mathrm{cat}} = \left[\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \boldsymbol{z}^{(3)}\right] \in \mathbb{R}^{384}.$$

以降の類似度計算に備え，各埋め込みは L2 正規化して用いる．

4.3.2　SCAdaCos とサブスペース損失

Wilkinghoff ら [37] に従い，各埋め込みには SCAdaCos 損失を適用し，クラス内分散の縮小とクラス間マージンの拡大を同時に図る．Fujimura ら [3] は，連結埋め込みと各ブランチ埋め込みの双方に SCAdaCos を課すことで，多解像度の一貫性と各サブ空間の識別性を両立させるサブスペース損失を導入した．その定義は

$$\mathcal{L}_{\mathrm{ss}} = \mathcal{L}_{\mathrm{SCAC}}\big(\boldsymbol{z}^{\mathrm{cat}}, \boldsymbol{l}\big) + \sum_{m=1}^{3} \mathcal{L}_{\mathrm{SCAC}}\big(\boldsymbol{z}^{(m)}, \boldsymbol{l}\big) \tag{4.1}$$

であり，ここで $\boldsymbol{l}$ は機械タイプと属性を結合したワンホットラベルベクトルである．第 1 項はグローバルなクラス中心を安定化させ，第 2 項は各サブ空間が単独でも識別的手掛かりを保持することを促す．データ拡張としては Mixup [68] を用いる．

4.3.3 推論

学習後，ソースドメインの埋め込みとターゲットドメインの埋め込みをそれぞれ k-means によりクラスタリングし，代表埋め込み集合を得る：

$$\mathcal{C}_{\mathrm{so}} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{k_{\mathrm{so}}}\}, \tag{4.2}$$

$$\mathcal{C}_{\mathrm{ta}} = \{\boldsymbol{c}_{k_{\mathrm{so}}+1}, \ldots, \boldsymbol{c}_{k_{\mathrm{so}}+k_{\mathrm{ta}}}\}, \tag{4.3}$$

$$\mathcal{C} = \mathcal{C}_{\mathrm{so}} \cup \mathcal{C}_{\mathrm{ta}}, \qquad J = k_{\mathrm{so}} + k_{\mathrm{ta}}. \tag{4.4}$$

ここで k_{so}, k_{ta} はそれぞれのドメインにおけるクラスタ数を表すハイパーパラメータである．テスト埋め込み $\boldsymbol{z}$ に対し，各代表埋め込みとのコサイン類似度

$$s_j(\boldsymbol{z}) = \frac{\langle \boldsymbol{z},\, \boldsymbol{c}_j \rangle}{\|\boldsymbol{z}\|\,\|\boldsymbol{c}_j\|}, \qquad j = 1, \ldots, J \tag{4.5}$$

を計算し，最大類似度の負値を異常スコアとする：

$$\mathrm{score}(\boldsymbol{z}) = -\max_{1 \le j \le J} s_j(\boldsymbol{z}). \tag{4.6}$$

スコアが大きいほど，$\boldsymbol{z}$ がいずれの正常クラスタからも離れていることを示す．

4.4 提案手法

ID や属性が利用できない未ラベル条件下では，識別モデルの性能は大きく低下することが知られている [3]．本研究ではサブスペース損失 $\mathcal{L}_{\mathrm{ss}}$ と推論手順に基づき，ラベル欠如の問題に対処し異常音検知性能を向上させる新たな手順を提案する．提案手法は次の 3 つの要素から構成される：

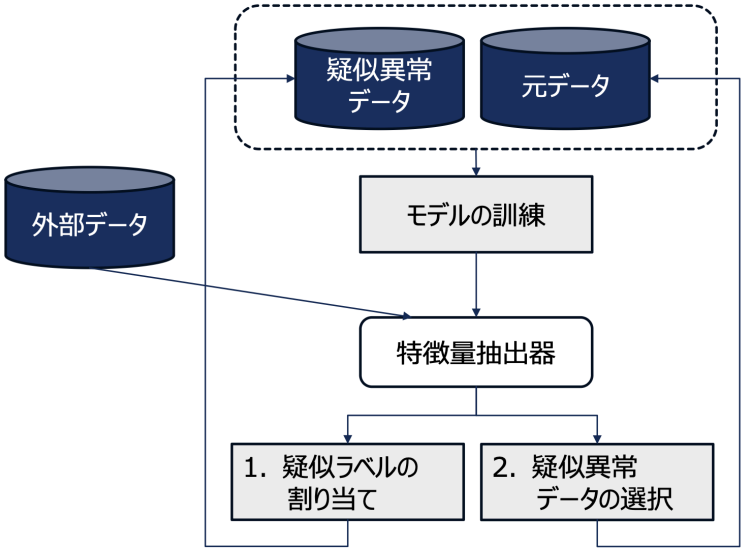


図 4.1: 提案手法の概要．反復学習フレームワーク内で統合された3つの主要構成要素：(1) 特徴量抽出器を用いた外部データからの擬似異常集合の選択，(2) 同じ特徴量抽出器による未ラベルの元データへの擬似ラベル付与，(3) 擬似異常集合と元データの両方から得られる更新済み学習データを用いてモデルを複数サイクル再学習し，性能を段階的に改善する反復学習．

1. **外部データからの擬似異常集合の選択：** 対象機械タイプの正常データに類似した外部データを，機械タイプごとの閾値に基づいて選択する．
2. **未ラベルデータへの擬似ラベル付与：** トリプレット学習を用いて未ラベルデータにクラスラベルを付与する．
3. **反復学習：** 性能を洗練するためにモデルを反復的に再学習する．

提案手法の概要を図 4.1 に示す．以下の小節で各要素の詳細を述べる．

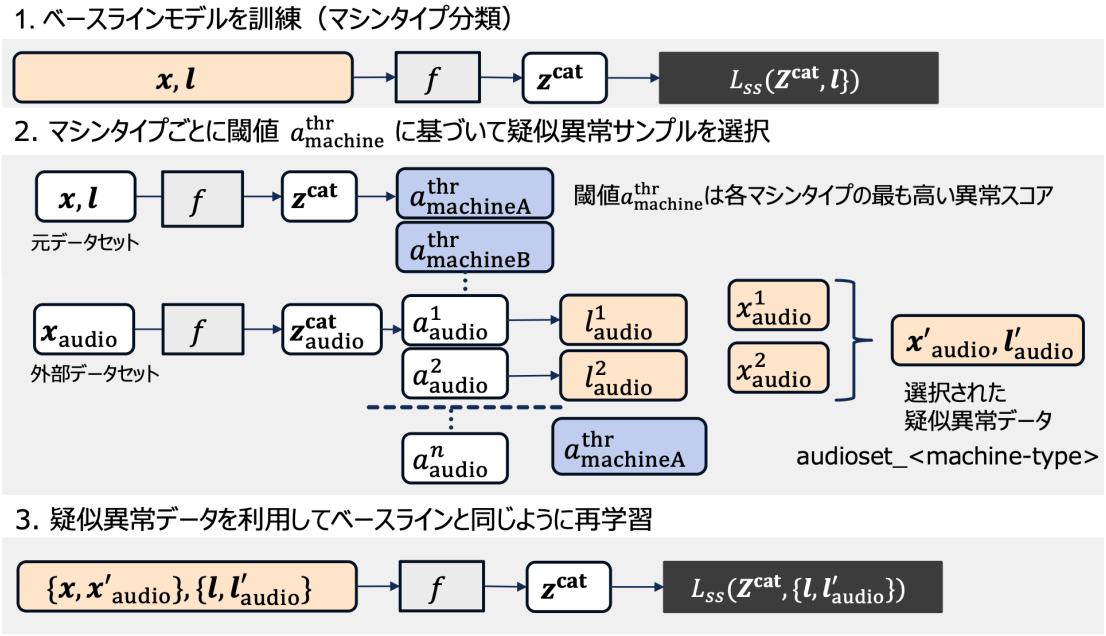


図 4.2: 外部データから擬似異常集合を選択する処理の概要．(1) サブスペース損失を用いて元データセット上でベースラインモデルを学習し，(2) 外部データをベースラインモデルに入力して異常スコアを算出し，機械タイプごとの閾値により選別して擬似異常集合を得て，(3) 得られた擬似異常集合を加えた結合データセットでベースラインモデルを再学習し，サブスペース損失により正常データの識別境界を洗練する．

4.4.1 外部データからの擬似異常集合の選択

同一機械タイプ内で属性ラベルなどの類似音を識別するためのラベルが存在しない場合，識別モデルの性能は低下する．提案法は，対象機械タイプの正常音に類似する外部データを選択し，適切なクラスラベルを付与することでこの問題に対処する．この処理の概要を図 4.2 に示す．擬似異常集合の選択は，(i) 機械タイプラベルを用いたベースラインモデルの学習，(ii) 機械タイプごとの閾値を用いた外部データからの擬似異常集合の選別，(iii) 擬似異常集合を加えたベースラインモデルの再学習，の三段階からなる．外部データ選別のためのモデル学習にはベースラインを用いる．

外部データからの擬似異常集合の選別

提案法ではベースラインモデルが正常と誤判定する外部データを選択する．学習済みベースラインモデルはすべての学習データに対して異常スコアを計算し，機械タイプごとにその最大値を閾値 $a^{\mathrm{thr}}_{\mathrm{machine}}$ とする．外部データのうち，異常スコアが $a^{\mathrm{thr}}_{\mathrm{machine}}$ 未満のものを，対応する機械タイプの正常データに類似しているとみなす．外部データへの過度な依存を避けるため，機械タイプごとに追加する外部サンプル数の上限を $N_{\max}$ に制限する．具体的には，選択された外部サンプル数を N_{out} とすると，追加するサンプル数は次式で与える：

$$N_{\mathrm{ex}} = \min(N_{\mathrm{out}}, N_{\max}), \tag{4.7}$$

ここで，異常スコアが小さい順に N_{ex} 個の外部サンプルを学習データに追加する．この拡張データセットを $\mathcal{X}_{\mathrm{train}}$ と呼ぶ．外部データの分類ラベルは `machine_attribute` 形式とし，以下の規則に従う：

- (`machine`)：元データセットの機械タイプに基づき，外部データセットに対する類似度が最大となるクラスを用いて割り当てる．
- (`attribute`)：外部データセット側にクラスラベルが存在する場合は，そのラベルを割り当てる．

同一の外部クラスに属するサンプルであっても，紐づく機械タイプが異なる場合は，外部データのクラス定義が粗い可能性を考慮し，別クラスとして扱う．

外部データを用いる従来法 [46], [90] では，外部データをランダムに選択して擬似異常と定義し，二値分類で学習することが多い．しかしその場合，楽器音や音声など異常音検知に無関係な音が選ばれる可能性があり，機械タイプによっては有効性が制限される．これに対し本手法は，異常音検知に有用な外部データのみをフィルタリングし，多クラス分類として扱うことで，正常データ間の微妙な差異を捉えられるようにし，性能を向上させる．

擬似異常集合を用いたモデルの再学習

特徴量抽出器は，拡張データセット $\mathcal{X}_{\text{train}}$ 上でベースラインと同じ手順により再学習する．学習後，代表埋め込みは外部データを除外し元データセットのみに基づいて算出する．これは，外部データに類似した異常音を正常と誤判定することを防ぐためである．異常スコアと代表埋め込みの計算方法はベースライン [3] と同一である．

4.4.2 未ラベルデータへの擬似ラベル付与

機械の内部状態や構成といった属性を表すラベルが利用できない場合の性能向上法を提案する．図 4.3 に本手法の概要を示す．処理は（i）ベースラインモデルの学習，（ii）ベースラインモデルによる擬似ラベルの取得，（iii）擬似ラベルを用いたベースラインモデルの再学習，の三段階からなる．

ベースラインモデルの学習

属性ラベルがない未ラベル条件では，ベースライン学習は通常，機械タイプラベルのみに依存する．しかしこれは分類課題を単純化しすぎるため，モデルが特定周波数や背景雑音といった無関係な特徴に着目してしまい [37]，微小な異常の検出能力を損なうことがある．この制約に対処し，効果的な擬似ラベル付与に備えるため，本研究では先行研究 [3] の示唆に基づきトリプレット学習を導入してベースライン学習を強化する．具体的に，[3] はトリプレット学習が運転音の変動と環境雑音を効果的に分離し，異常検知に適した埋め込み空間を形成できることを示している．本研究では，異なるサンプル間の分離よりも，同一サンプル内の微妙な変化（機械の運転音の変化など）を捉えることが目的により適合すると考え，コントラスト学習ではなくトリプレット学習を選択した．コントラスト学習はミニバッチ内の異なるサンプルを分離することに焦点を当てるため，微細差の識別には相対的に不利である．これに対しトリプレッ

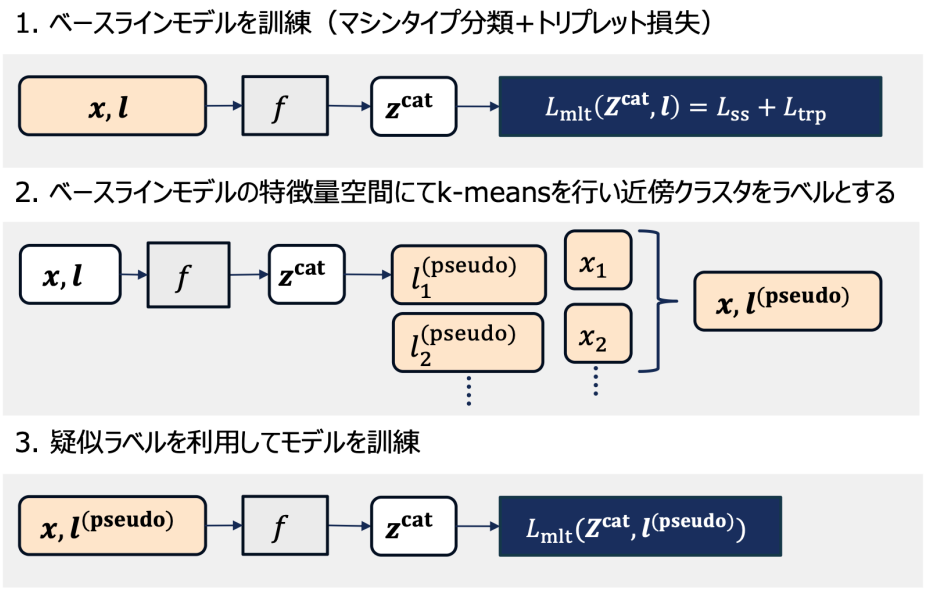


図 4.3: 未ラベルデータに擬似ラベルを付与する提案法の概要．(1) 元データセットを特徴量抽出器 f に通してカテゴリ出力 $\boldsymbol{z}^{\mathrm{cat}}$ を得るベースラインモデルを学習し，損失関数 $\mathcal{L}_{\mathrm{mlt}}$ で最適化，(2) 学習済みベースラインモデルで未ラベルデータの予測を行い擬似ラベル $\boldsymbol{x}, \ell^{(\mathrm{pseudo})}$ を生成，(3) 擬似ラベル付きデータ $\boldsymbol{x}, \ell^{(\mathrm{pseudo})}$ を用いてモデルを再学習し，損失関数により予測を反復的に改善する．

ト学習は，無関係な雑音変動を無視しつつ，これらの微妙な「同一サンプル内の変化」を強調するようモデルを促す．

トリプレット学習では，アンカー $\boldsymbol{x}_i^a$，ポジティブ $\boldsymbol{x}_i^p$，ネガティブ $\boldsymbol{x}_i^n$ を次のように定義する：

- **アンカー**：アンカー $\boldsymbol{x}_i^a$ は (i) 番目の機械タイプの正常音サンプル．
- **ポジティブ**：ポジティブ $\boldsymbol{x}_i^p$ は，別機械タイプ $j \neq i$ の音を背景雑音としてスケーリング加算して生成する：

$$\boldsymbol{x}_i^p = \boldsymbol{x}_i^a + 10^{-\frac{\alpha}{20}} \cdot \frac{|\boldsymbol{x}_i^a|}{|\boldsymbol{x}_j|} \boldsymbol{x}_j, \tag{4.8}$$

ここで $|\cdot|$ はユークリッドノルム，α は信号対雑音比（SNR）を dB で表すハイパーパラメータであり，アンカーに対する背景雑音の強度を調整する．

- **ネガティブ**：ネガティブ $\boldsymbol{x}_i^n$ はアンカーにピッチシフトを施して生成する：

$$\boldsymbol{x}_i^n = \mathrm{PitchShift}(\boldsymbol{x}_i^a, \beta), \tag{4.9}$$

ここで β は運転状態の変化を模擬し，実装は torchaudio [128], [129] に基づく．

このトリプレット構成により，モデルは背景雑音の変動と運転状態の変化を識別するよう促される．$\boldsymbol{z}_i^a$，$\boldsymbol{z}_i^p$，$\boldsymbol{z}_i^n$ をそれぞれ $\boldsymbol{x}_i^a$，$\boldsymbol{x}_i^p$，$\boldsymbol{x}_i^n$ の埋め込みとする．

温度パラメータ τ をもつ類似度関数を用いる：

$$s_\tau(\boldsymbol{z}, \boldsymbol{z}') = \frac{\langle \boldsymbol{z}, \boldsymbol{z}' \rangle}{\tau \, |\boldsymbol{z}||\boldsymbol{z}'|}, \tag{4.10}$$

ここで $\langle \cdot, \cdot \rangle$ は内積を表す．トリプレット損失 $\mathcal{L}_{\mathrm{trp}}$ は次式で定義する：

$$\mathcal{L}_{\mathrm{trp}}(\boldsymbol{z}_i^a, \boldsymbol{z}_i^p, \boldsymbol{z}_i^n) = \max\left\{0, \gamma + 1 - s_\tau(\boldsymbol{z}_i^a, \boldsymbol{z}_i^p) + s_\tau(\boldsymbol{z}_i^a, \boldsymbol{z}_i^n)\right\}, \tag{4.11}$$

ここで γ はマージンである．この損失は，モデルが雑音ではなく運転音の差異を優先して捉えることを促す．

特徴量抽出器は，トリプレット損失 $\mathcal{L}_{\mathrm{trp}}$ とサブスペース損失 $\mathcal{L}_{\mathrm{ss}}$ の双方で学習する．ここで $\boldsymbol{l}_i$ は機械タイプのワンホットラベルである．$\mathcal{L}_{\mathrm{ss}}$ には 50 % の確率で Mixup [68] を適用するが，$\mathcal{L}_{\mathrm{trp}}$ にはトリプレット関係を保持するため適用しない．最終的な学習損失は

$$\mathcal{L}_{\mathrm{mlt}} = \mathcal{L}_{\mathrm{trp}} + \lambda \, \mathcal{L}_{\mathrm{ss}}, \tag{4.12}$$

である．ここで λ は両損失のバランスを決めるハイパーパラメータである．この強化されたベースライン学習により堅牢な埋め込み空間が形成され，後続の擬似ラベル付与が有効となり，反復学習フレームワークを支える．

擬似ラベルの生成

学習後，各ドメインかつ各機械タイプごとに，得られた埋め込みに対して k-means クラスタリングを適用して擬似ラベルを生成する．これは先行研究 [3], [130], [131] と同様である．ソースドメインとターゲットドメインのクラスタ数をそれぞれ k_{so} と k_{ta} とする．各サンプル $\boldsymbol{x}_i$ は埋め込み空間において最も近いクラスタに割り当てる：

$$\ell_i^{(\mathrm{pseudo})} = \underset{1\le j\le k_{d_i}}{\arg\min} \left\| \boldsymbol{z}_i - \boldsymbol{c}_j^{(d_i)} \right\|, \quad \text{where } d_i = \begin{cases} \mathrm{so} & (\text{source domain}), \\ \mathrm{ta} & (\text{target domain}). \end{cases} \tag{4.13}$$

擬似ラベルを用いたベースラインモデルの再学習

$\boldsymbol{l}_i$ を擬似ラベル $\ell_i^{(\mathrm{pseudo})}$ に置き換え，$\mathcal{L}_{\mathrm{mlt}}$ を用いてモデルをゼロから再学習する．擬似ラベルは真の教師信号ではないため，$\mathcal{L}_{\mathrm{ss}}$ のみによりクラス内分散を縮小すると，真のラベルが異なるサンプルが同一クラスタに押し込められ，環境雑音に着目したり運転音を誤って群化したりする恐れがある．トリプレット損失を併用することで，モデルは雑音の違いを無視し運転音の変化に焦点を当てて学習でき，これらの問題を緩和して異常音検知性能を向上させる．推論時の手順はベースラインと同一である．

4.4.3　擬似異常集合と擬似ラベルの反復的選択

擬似異常集合選択と擬似ラベル付与を，次の反復学習スキームに統合する：

- **Stage 1**：外部データや擬似ラベルを用いず，元のラベル付きデータセットに対して $\mathcal{L}_{\mathrm{mlt}}$ でモデルを学習する．
- **Stage** M（$M \ge 2$）：
  1. Stage $(M-1)$ のモデルを用いて擬似異常集合を選択し，機械タイプごとに最大 $N_{\max}$ 個まで学習集合に追加する．

2. 同じモデルを用いて k-means により擬似ラベルを付与する．

3. 拡張データセットと新たな擬似ラベルを用い，$\mathcal{L}_{\mathrm{mlt}}$ でモデルを再学習する．

本手法の概要は図 4.1 に再掲のとおりである．推論時には，ベースラインと同様に，代表埋め込みとの最大類似度に基づいて異常スコアを計算する．この反復的アプローチは，外部データと擬似ラベル化された内部変動の双方を活用することで，各段階でモデルを洗練し，性能を段階的に向上させる．

4.5 実験的評価

4.5.1 データセット

本研究では，2022–2024 年の DCASE Task 2 データセット [10], [11], [12] を用いて提案手法を評価した．表 4.1 にデータセットの詳細を示した．これらのデータセットは工場の背景雑音を含む機械音録音から構成される．各録音は単一チャネルの音声ファイルであり, 長さは 6–18 秒, サンプリング周波数は 16 kHz である．DCASE2022 データセットには 7 種類の機械タイプ（fan, gearbox, bearing, slide rail（slider）, valve, ToyCar, ToyTrain）が含まれる [111], [132]．DCASE2023 データセットは 14 種類の機械タイプで構成され，開発セットは DCASE2022 と同一である一方，評価セットには ToyDrone, ToyTank, ToyNscale, bandsaw, grinder, shaker が含まれる [133]．DCASE2024 データセットは 16 種類の機械タイプで構成され，開発セットは DCASE2022 に整合し，評価セットには 3D-Printer, AirCompressor, BrushlessMotor, HairDryer, HoveringDrone, RoboticArm, Scanner, ToothBrush, ToyCircuit が含まれる [134], [135]．各データセットのラベル条件を表 4.1 に要約する．なお，DCASE2024 の一部の属性ラベルは競技中には利用できなかったが，競技終了後に公開され，本解析で使用した．

各データセットは機械タイプごとに正常データを 1,000 個の学習サンプル提供する．これらは 990 個のソースドメインサンプルと，ドメインシフトの影響を受けた 10 個の

表 4.1: 各年度における DCASE Task 2 にて使用されたデータセットの詳細.

DCASE	# 機械タイプ	# ID	# 機械タイプあたりの訓練データ
2022	7	6	6000
2023	14	1	1000
2024	16	1	1000

ターゲットドメインサンプルから構成される．ドメインシフトは，保守作業による機械音特性の変化や，背景雑音や運転条件の変動などの音響環境の違いに起因して発生する．学習サンプルには，機械の運転状態や環境を示す属性ラベルが含まれる．理想的な異常音検知システムは適応を行わずともドメインシフトに頑健に異常を検出できるべきである [51]．DCASE Task 2 の設定に従い，学習データにはソース / ターゲットのドメイン情報が示される．評価データは機械タイプごとに 100 個の正常サンプルと 100 個の異常サンプルを含み，両ドメインで均等に分割される．推論時にはテストデータのドメインは未知である．DCASE2022 では ID がドメインシフトの種類を示し，ラベルあり設定では属性ラベルと併用される．DCASE2023 以降は ID が存在せず，分類課題が単純化された結果，異常変化に敏感な埋め込みの抽出が難しくなった．性能評価には DCASE Task 2 に従って AUC を用いる [136], [137].

4.5.2　システム記述

本研究では，未ラベル設定の提案法を，従来異常音検知システムで複数データセットにわたり最先端性能を示す Wilkinghoff 法 [125] と，Fujimura らによるベースライン [3] と比較する．異常音検知では異常データを用いたハイパーパラメータ調整ができないため，機械タイプをまたいで同一のハイパーパラメータを用いる．このアプローチは異常音検知で広く採用されており [10], [11], [12]，未知の機械タイプに対しても堅

牢な性能を保証する．各手法の設定を以下に示す．

Wilkinghoff 法（Wilkinghoff [125]）．本手法は [125] に従い，入力音響特徴として振幅スペクトログラムと全振幅スペクトルを用いる．入力音響特徴の抽出では窓長 64 ms，ホップサイズ 50 % を用いる．2 本の畳み込みブランチがそれぞれ 128 次元の埋め込みを生成し，これらを連結して 256 次元の埋め込みとして異常音検知に用いる．SCAdaCos 損失 [37] を，この 256 次元の埋め込みに対して，ランダム初期化の学習不可な 16 個のサブクラスタに適用する．学習はバッチサイズ 100，50 エポック，学習率 0.001 の AdamW 最適化法 [119]，および一様サンプリング比率による Mixup を用いた．学習後，埋め込みを k-means でクラスタリングし，ソースドメインに 16 クラスタ，ターゲットドメインに 10 クラスタを設定し，ターゲットドメインでは各サンプルがそのまま代表埋め込みとして用いられる．DCASE2023 Task 2 における最適なソースドメインのクラスタ数は 16 であったが，クラスタ数の変化による影響は小さかった [125]．異常スコアはテストサンプルと代表埋め込みとのコサイン類似度から算出する．

ベースライン（Ba [3]）．このベースラインは Wilkinghoff [125] の設定を多く踏襲するが，主な相違は次のとおりである：(i) 8 ms と 256 ms の 2 種類の振幅スペクトログラムに全振幅スペクトルを加え，いずれもホップサイズ 50 %を用いる．(ii) 3 本の畳み込みブランチがそれぞれ 128 次元埋め込みを出力し，連結して 384 次元の埋め込みを構成；(iii) SCAdaCos 損失 [37] を 2 度適用し，まず 384 次元の埋め込みに対して学習不可な 16 サブクラスタ，ついで各 128 次元ブランチごとにランダムに初期化された学習可能な 16 サブクラスタを用いる．その他の学習・推論パラメータは Wilkinghoff 法に整合する．

ベースライン＋擬似異常データ選択（Ba+Ex）. 外部データとして Audioset [138] の約 180 万件のラベル付き音声サンプルを取り込む．各 Audioset サンプルには「mid」ラベル（例，ピアノ，男性音声, 機械の動作音）が付与されており，これを属性ラベル

として用いる．代表埋め込みは学習集合から k-means により算出し，ソースドメインには 16 クラスタを設定する．それぞれの機械タイプの選択可能な擬似異常データの上限は学習データと同数とするので最大 $N_{\max} = 1000$ 個まで外部サンプルを選択可能とする．ベースラインモデルは，これらのラベル付きサンプルを追加して再学習し，元のハイパーパラメータを維持する．

ベースライン＋トリプレット損失（$\mathtt{Ba} + \mathcal{L}_{\mathbf{trp}}$）. ベースラインにトリプレット損失を導入して強化する．ポジティブサンプルは，別機械タイプの音を背景雑音としてスケーリング加算して生成し，ハイパーパラメータ α（SNR $[-5, 20]$ dB）で雑音強度を制御する．SNR の範囲は，雑音と原音の強度バランスをとり，両者が同程度に比較可能となるように選定した．ネガティブサンプルはアンカーにピッチシフトを適用して生成し，実装は torchaudio [128], [129] に基づく．β の範囲は，± 6〜± 12 とし原音を過度に歪ませず現実的な運転変動を模擬できるよう，半音から 1 オクターブまでをカバーするように選択した．トリプレット損失の設定は [139] に従い，$\tau = 0.2$，$\gamma = 0.5$ を用いる．

ベースライン＋擬似ラベル（Ba+Ps）. 初期学習後，各ドメインごとに埋め込みをクラスタリングし，ソースドメインは 16 クラスタ（$k_{\mathrm{so}} = 16$），ターゲットドメインは機械タイプごとに 4 クラスタ（$k_{\mathrm{ta}} = 4$）とする．ソースに 16 属性，ターゲットに 4 属性があると仮定すると，機械タイプあたり 20 の擬似クラスとなる．サンプルは 20 個のクラスタの中心のうち最近傍に基づいて擬似ラベルに割り当てられる．Stage 2 以降，モデルは Stage 1 の機械タイプ数に対して 20 倍のクラス数を持つ分類課題に取り組むことになる．

反復学習法（$\mathtt{Ba}+\mathcal{L}_{\mathbf{trp}}+\mathtt{Ps}+\mathtt{Ex}$, Stage M）. 本反復法は最大 $M = 5$ 回の反復で性能を高める．Stage $M = 2, 3, 4, 5$ では，Stage $M - 1$ のモデルを用いて擬似ラベルと外部データを導出する．

表 4.2: DCASE 2022–2024 の Task 2 データセットに対する各学習設定の AUC [%] の平均値．上段（“w/ label”）はラベルを用いた参照結果，下段（“w/o label”）は提案する未ラベル設定の結果を示す．“stage” 列は反復回数を表し，stage 1 は初期モデル，stage 2 は stage 1 から得た外部データまたは擬似ラベルを用いて更新し，stage 3–5 は同じ更新手順を再帰的に繰り返す．“dev” と “eval” はそれぞれ開発・評価セットに対応する．値は 5 つの乱数シードを用いて計算した平均と標準偏差を表す．`Ba` = baseline [3]，`Ex` = 選択した外部データ，`Ps` = 擬似ラベル，$\mathcal{L}_{\text{trp}}$ = トリプレット損失を表す．

Use label	Method	stage	2022		2023		2024	
			dev	eval	dev	eval	dev	eval
w/ label	Wilkinghoff [125]	1	82.5±0.8	73.1±0.9	67.2±0.8	74.2±0.3	72.6±0.7	61.5±0.6
	`Ba` [3]	1	81.9±0.9	73.0±0.3	70.5±0.5	77.5±0.4	72.8±0.4	63.2±0.8
w/o label	Wilkinghoff [125]	1	67.2±5.8	64.1±0.8	62.5±0.8	64.4±3.2	59.7±1.1	55.6±0.6
	`Ba` [3]	1	71.3±0.9	64.8±0.7	64.2±1.2	67.8±1.4	59.5±0.7	53.8±0.6
	`Ba`+$\mathcal{L}_{\text{trp}}$	1	71.8±1.6	65.2±1.3	64.1±1.2	68.8±0.6	59.7±1.3	54.1±1.5
	`Ba`+$\mathcal{L}_{\text{trp}}$+`Ex`	2	74.0±0.5	65.3±1.3	65.0±1.1	69.2±0.3	61.2±1.3	54.9±0.9
	`Ba`+$\mathcal{L}_{\text{trp}}$+`Ps`	2	76.2±0.4	68.4±0.8	64.2±1.3	72.4±0.7	65.5±1.5	56.4±1.1
	`Ba`+$\mathcal{L}_{\text{trp}}$+`Ps`+`Ex`	2	75.8±0.9	69.1±0.7	64.4±0.5	72.7±1.0	66.4±2.2	56.5±0.6
	`Ba`+$\mathcal{L}_{\text{trp}}$+`Ps`+`Ex`	3	76.8±1.7	70.1±1.4	**65.2±0.5**	72.6±1.3	68.3±1.0	**57.0±0.1**
	`Ba`+$\mathcal{L}_{\text{trp}}$+`Ps`+`Ex`	4	76.5±2.0	70.1±0.6	64.3±1.9	**73.1±0.6**	**70.7±1.5**	56.1±0.0
	`Ba`+$\mathcal{L}_{\text{trp}}$+`Ps`+`Ex`	5	**78.1±1.0**	**70.3±1.1**	**65.2±1.4**	72.6±0.3	70.4±1.5	56.8±0.1

4.5.3 ラベル無し条件での評価

表 4.2 は，未ラベル条件下で提案手法の性能を評価した実験結果を示す．ドメイン別の性能は付録の 表 5.2 に示す．まず $\mathcal{L}_{\text{trp}}$ の効果を検証するため，`Ba` と `Ba`+$\mathcal{L}_{\text{trp}}$ を比較した．$\mathcal{L}_{\text{trp}}$ を導入することで，DCASE2024 の開発セットを除く全データセットで性能が向上し，その差も僅少であった．これは，$\mathcal{L}_{\text{trp}}$ が一般に有効であり，サンプル内の運転音変動を強調しつつ雑音を抑制することで，より識別的な埋め込みを形成していることを示唆する．

次に，外部データで再学習する効果を検証するため，Ba+$\mathcal{L}_{\mathrm{trp}}$ と Ba+$\mathcal{L}_{\mathrm{trp}}$+Ex を比較した．外部データの利用により全データセットで性能が向上し，その有効性が示された．特に，正常データに類似した外部データを追加することで，それら外部サンプルに近い異常の検出能力が高まったと考えられる．

さらに，擬似ラベルで再学習する効果を確認するため，Ba+$\mathcal{L}_{\mathrm{trp}}$ と Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps を比較した．擬似ラベルを用いた再学習は全データセットで性能を改善し，その有効性が示された．運転音が類似するサンプルを同一クラスとして扱うことで，機械設定の違いといったより微細な差異も検出しやすくなり，性能向上につながったと考えられる．

外部データと擬似ラベルの併用効果を検討するため，Ba+$\mathcal{L}_{\mathrm{trp}}$+Ex，Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps，Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex を比較した．多くのデータセットで併用法が個別法を上回り，例外は DCASE2022 の開発セット（Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps が最良）と DCASE2024 の開発セット（Ba+$\mathcal{L}_{\mathrm{trp}}$+Ex が最良）のみであった．これらの例外ではターゲットドメインで特に高い性能が得られており，ターゲットに適した外部データや擬似ラベルが取得できた可能性がある．いずれのデータセットにおいても，併用が各単独手法に劣化する例は見られなかったため，両者の併用は有効と考えられる．

反復学習による性能向上を評価するため，Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex の stage 2，3，4，5 を比較した．全データセット平均は stage 2 の 67.5 から，stage 3，4，5 でそれぞれ 68.3，68.5，68.9 と着実に向上した．ソース・ターゲットの両ドメインで性能が改善し，提案法が未ラベル条件で有効に性能を底上げすることが示された．個別データセットでも，stage 3–5 は概ね stage 2 を上回ったが，stage 3–5 間の差は顕著ではなかった．異常データが検証に利用できない前提を踏まえると，未知機械向けの異常音検知構築では stage 3 まで反復することが実務上妥当と考えられる．

最後に，真の属性ラベルを用いた上界性能と未ラベル設定の結果を比較した．すべてのデータセットの平均は，w/ label の Wilkinghoff [125] が 71.9，w/ label の Ba [3] が 73.2，w/o label の Wilkinghoff [125] が 62.3，w/o label の Ba [3] が 63.6 であった．

Ba [3] におけるラベル有無の差は 9.4 ポイントであったが，本手法の stage 5 ではこの差を 4.3 ポイントまで縮小し，5.1 ポイント改善した．さらに，stage 5 は w/ label の Wilkinghoff [125] に対しても 3.0 ポイント差まで迫った．これらより，提案法は既存の未ラベル手法を大きく上回り，ラベルありの性能に近づけることが確認できた．

以上の結果から，未ラベル設定では $\mathcal{L}_{\text{trp}}$・外部データ・擬似ラベルを併用し，少なくとも 3 回の反復を行うことが有効である．

4.5.4 ラベル有り条件での評価

表 4.3 に示した結果を用いて，ラベルあり設定において提案手法を評価した．ドメイン別の性能は付録の 表 5.3 に示す．擬似ラベル手法は，元のラベルが異常音検知に対して十分な細粒度を欠いている場合を仮定し，これを補間するために活用する．本手法は，より細かな分類を可能にするために，属性ラベルに追加で擬似ラベルを付加する．具体的には，擬似ラベルを適用する場合，ラベル形式は `machine_attribute` から `machine_attribute_pseudo-label` へと移行する．

トリプレット学習の影響を評価するために，Ba [3] と Ba+$\mathcal{L}_{\text{trp}}$ を比較した．真の属性ラベルが利用可能な場合，$\mathcal{L}_{\text{trp}}$ を導入すると，全データセットで性能が低下する．$\mathcal{L}_{\text{trp}}$ はサンプル内の微細な変動を捉えるようモデルを駆動するため，元ラベルがすでに機械設定を適切に表現している状況では，微小な音の差を過度に強調してしまう可能性がある．この過敏さが性能低下の主因であると考えられる．

次に，Ba [3] と Ba+Ex を比較して外部データの効果を検証した．この設定では $\mathcal{L}_{\text{trp}}$ が性能を向上させないため，stage 1 のモデルとして Ba [3] を用いる．外部データの追加により，DCASE2023 および DCASE2024 のすべてのデータセットで性能が向上した．一方で，DCASE2022 では有意な改善は見られなかった．外部データを用いる場合，正常データに類似するサンプルが外部ソースから選択される．機械タイプの区別があらかじめ定義されていない場合や，類似した機械音が元のデータセット内に存在

表 4.3: DCASE 2022–2024 の Task 2 データセットに対する，各教師あり設定（属性ラベル利用可）の AUC [%] の平均値．“stage” 列は学習の反復を表し，stage 1 は初期モデル，stage 2 は stage 1 から得た外部データ（Ex）および擬似ラベル（Ps）でベースラインを再学習，stage 3 は同手順を繰り返す．“dev” と “eval” はそれぞれ開発・評価セットを表し，値は 5 つの乱数シードを用いて計算した平均と標準偏差を表す．`Ba` = baseline [3]，`Ex` = 選択した外部データ，`Ps` = 擬似ラベル，$\mathcal{L}_{\text{trp}}$ = トリプレット損失を表す．

Method	stage	2022		2023		2024	
		dev	eval	dev	eval	dev	eval
Wilkinghoff [125]	1	**82.5±0.8**	74.2±0.3	73.1±0.9	72.6±0.7	67.2±0.8	61.5±0.6
`Ba` [3]	1	81.9±0.9	**77.5±0.4**	73.0±0.3	72.8±0.4	70.5±0.5	63.2±0.8
`Ba`+$\mathcal{L}_{\text{trp}}$	1	80.7±0.8	68.8±0.3	72.6±2.3	69.0±0.9	68.8±0.9	61.9±1.6
`Ba+Ex`	2	81.7±0.3	76.9±0.5	**73.8±1.0**	**75.0±0.9**	**71.2±0.9**	**64.4±0.6**
`Ba+Ps`	2	79.8±0.4	76.0±0.4	73.6±0.6	71.7±1.6	70.2±0.6	61.0±1.1
`Ba+Ps`+$\mathcal{L}_{\text{trp}}$	2	80.0±0.5	75.8±0.5	72.5±0.6	69.0±1.6	69.1±0.9	60.7±0.8
`Ba+Ex`	3	82.2±0.8	76.8±0.5	73.4±0.5	**75.0±1.5**	70.2±1.1	**64.4±1.2**

しない場合には，外部データが分類の難易度を高め，性能向上につながる可能性がある．これに対し，ID が存在する状況では，学習データ自体が既に類似した運転音を十分に含んでいるため，外部データによるタスク難易度の上昇効果は限定的で，得られる利得も小さいと考えられる．

続いて，`Ba` [3] と `Ba+Ps` を比較し，擬似ラベリングの効果を評価した．擬似ラベルは，DCASE2023 の開発セットを除くすべてのデータセットで性能を低下させた．ラベルあり設定では，擬似ラベリングは既存ラベルをより細かなカテゴリへと分割する働きを持つ．もし元ラベルが機械設定を適切に反映している場合，さらなる細分化はデータを不必要に分割してしまう．こうして追加されたカテゴリは背景雑音などの無関係な変動を捉えてしまう可能性があり，元データが適切にセグメント化されている

場合には擬似ラベリングが冗長となり得る．

最後に，反復学習の利点を評価するために，Ba+Ex の stage 2 と stage 3 を比較した．この設定では $\mathcal{L}_{\mathrm{trp}}$ と擬似ラベリングのいずれも有用でないため，Ba+Ex のみで反復を行う．結果として，stage 2 から stage 3 への改善はほとんど見られなかった．stage 2 では，stage 1 の選別に基づいて，モデルは Audioset のサンプルと正常データとの識別を学習する．$\mathcal{L}_{\mathrm{trp}}$ や擬似ラベリングから新たな知見が得られない状況では，Audioset から外部データを再抽出しても追加の観点は得られず，性能はほぼ変わらない．

これらの結果は，元ラベルが適切にアノテーションされているラベルあり設定においては，外部データ手法を 1 回適用することが最も有効であることを示唆している．

4.5.5 外部データ選択の有効性と外部データ量の影響

未ラベル設定において，外部データに由来する擬似異常データの効果を調査した．表 4.4 は以下の 3 点に着目して解析した：

1. 異常スコアによる外部データ選択（提案法）とランダム選択の性能差,
2. 学習に用いる外部サンプルの最大数 N_{max} による性能の変動,
3. $N_{\mathrm{out}} \geq N_{\mathrm{max}}$ の機械タイプ（*large* N_{out} *machines*）と $N_{\mathrm{out}} < N_{\mathrm{max}}$ の機械タイプ（*small* N_{out} *machines*）の性能差.

DCASE2022 では，*large* N_{out} *machines* は fan と valve，*small* N_{out} *machines* は bearing，gearbox，slider，ToyCar，ToyTrain である．DCASE2023 では，*large* N_{out} *machines* は bandsaw と grinder，*small* N_{out} *machines* は bearing，fan，gearbox，shaker, slider, ToyCar, ToyDrone, ToyNscale, ToyTank, ToyTrain, Vacuum, valve を含む．DCASE2024 では，*large* N_{out} *machines* は BrushlessMotor, *small* N_{out} *machines* は 3DPrinter，AirCompressor，bearing，fan，gearbox，HairDryer，HoveringDrone，RoboticArm，Scanner，slider，ToothBrush，ToyCar，ToyCircuit，ToyTrain，valve で

表 4.4: 異常スコアによる外部データ選択とランダム選択を, 外部データの最大数（$N_{\max}$）を変化させて比較した性能評価．*Large* N_{out} *machines* は $N_{\text{out}} \geq N_{\max}$ の機械タイプを，*small* N_{out} *machines* は $N_{\text{out}} < N_{\max}$ の機械タイプを指す．各値は当該データセット内の全機械タイプにわたる AUC［%］の平均値を 5 つの乱数シードを用いて計算した平均と標準偏差を表す．

	$N_{\max}$	*large* N_{out} *machines*			*small* N_{out} *machines*		
		2022	2023	2024	2022	2023	2024
Ba [3]	–	71.8±1.2	58.4±0.6	59.9±2.3	67.3±0.7	65.6±0.5	59.2±0.6
Ba+Ex	500	74.1±1.2	61.2±1.9	65.1±2.7	68.3±0.6	**66.6±0.8**	59.3±0.8
Ba+Ex	1000	**74.6±0.8**	**62.3±2.2**	**66.1±5.7**	68.7±1.1	66.3±1.5	**60.0±0.8**
Ba+Ex	2000	74.1±0.9	60.6±2.5	63.8±5.4	**69.0±0.6**	66.3±0.6	**60.0±1.0**
Ba+Ex (random)	500	73.3±0.9	60.3±2.4	63.6±2.4	67.8±0.9	66.1±1.2	59.8±0.8
Ba+Ex (random)	1000	73.5±1.1	60.6±1.9	59.2±2.9	67.4±0.3	65.8±1.0	59.3±0.6
Ba+Ex (random)	2000	74.1±0.7	61.5±1.5	61.1±1.1	67.6±0.5	65.1±0.5	58.9±1.5

ある．

提案法とランダム選択を比較するため, Ba [3] と Ba+Ex を評価した．Ba+Ex は $N_{\max}$ の値にかかわらず一貫して性能を向上させたのに対し，Ba+Ex（random）はデータセットによっては Ba [3] を下回る場合が見られた．特筆すべきは，いずれのデータセットにおいてもランダム選択ではなく提案法で最高性能が得られた点である．これは，正常と誤判定されやすい外部サンプルを的確に選別して取り込むことが，ランダム選択よりも効果的に異常音検知性能を高めることを示唆している．さらに，この改善は *large* N_{out} *machines* と *small* N_{out} *machines* のいずれに対しても観測され，ターゲット機械タイプだけでなく，同時学習している他機械タイプに類似した外部データを加えることも性能向上に寄与することを示している．

次に，Ba+Ex における $N_{\max}$ と性能の関係を検討した．*large* N_{out} *machines* では，$N_{\max}$ を 1000 から 500 あるいは 2000 に変更すると性能が低下した．これは，正常と

誤判定されやすい外部データを増やすこと自体は有効である一方で，元の学習データ量を超えて過剰に外部データを導入すると性能劣化を招くことを示す．すなわち，外部データ量と元データ量のバランスが重要である．一方，*small* N_{out} *machines* では，$N_{\max}$ を変えても一貫した傾向は見られなかった．$N_{\mathrm{out}} < N_{\max}$ のため，$N_{\max}$ を増やしても，ターゲットクラスではなく他機械タイプに類似した外部データが追加されることになり，性能に明確な変化が現れにくいと考えられる．

以上の結果から，正常と誤判定されやすい外部データを優先して選択する本手法は，ランダム選択よりも高い有効性を示し，特に関連する外部データが豊富な機械タイプ（*large* N_{out} *machines*）で性能を効果的に向上させることが分かった．さらに，これらの機械タイプに対しては，学習データ規模と同程度の $N_{\max} = 1000$ に設定することが重要であり，外部データの過剰導入による性能劣化を防ぐうえで有効である．

4.5.6 トリプレット損失と疑似ラベルの性能分析

表 4.5 は，stage 1 において Ba または Ba+$\mathcal{L}_{\mathrm{trp}}$ により生成した擬似ラベルを用い，stage 2 で学習したモデルの性能を評価している．結果は，stage 1 で Ba+$\mathcal{L}_{\mathrm{trp}}$ により生成された擬似ラベルのほうが，stage 2 の損失関数設定に依らず，Ba のみで生成した擬似ラベルより一貫して高い性能をもたらすことを示している．これは，stage 1 に $\mathcal{L}_{\mathrm{trp}}$ を導入することで擬似ラベルの品質が向上することを示唆している．

さらに，stage 2 の構成を比較すると，Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps は，全データセットかついずれの擬似ラベル条件でも Ba+Ps を上回った．このことは，stage 2 に $\mathcal{L}_{\mathrm{trp}}$ を追加することで，誤って割り当てられた擬似ラベルの負の影響が低減されることを示している．トリプレット損失は，雑音や機械固有の些末な特徴を無視しつつ，機械音の変化といった異常音検知に有用な特徴に着目するようモデルを訓練することで，擬似ラベルの品質向上と誤分類の抑制の双方に寄与する．

以上の知見から，擬似ラベルを用いる教師なし学習フレームワークにおいては，す

表 4.5: 二段階学習フレームワークにおいて，トリプレット損失 $\mathcal{L}_{\text{trp}}$ が擬似ラベルの品質およびモデル性能に与える影響の評価．列は，stage 1 で擬似ラベル生成に用いたモデル（Ba，Ba + $\mathcal{L}_{\text{trp}}$）とデータセット（DCASE2022，DCASE2023，DCASE2024）を表す．行は，stage 2 におけるモデル構成（Ba + Ps，Ba + $\mathcal{L}_{\text{trp}}$ + Ps）を示す．各セルは当該データセットの全機械タイプにわたる AUC [%] の平均値であり，5 回の異なるシードによる実行から平均と標準偏差を算出した．

stage 1 の損失関数	Ba			Ba+$\mathcal{L}_{\text{trp}}$		
DCASE	2022	2023	2024	2022	2023	2024
Ba+Ps	70.6±0.2	63.2±1.1	57.4±1.2	71.3±0.5	63.5±1.3	58.9±0.6
Ba+$\mathcal{L}_{\text{trp}}$+Ps	**71.8±0.4**	**66.8±1.4**	**59.1±0.5**	**74.3±0.5**	**67.0±1.0**	**60.3±1.2**

べての段階でトリプレット損失を取り入れることが有効であり，モデル性能の向上につながることが分かった．

4.6　結論

本研究では，詳細な運転状態ラベルや類似機械データが限られる状況において異常音検知性能を高めるための 3 つの手法を提案した．第一に，比較可能な機械タイプが乏しい状況に対処するための擬似異常集合選択法を提案した．大規模な外部データセットを走査し，対象機械の正常音に類似した音声サンプルを自動抽出することで，重要な特性を損なうことなく分類の難易度を高めた．第二に，未ラベルデータに対する擬似ラベル付与戦略を開発し，異常音検知に不可欠な微細な運転差を検出可能にした．学習済み埋め込みのクラスタリングにより未ラベルデータを擬似クラスへ細分化し，微粒度な異常に焦点を当てるようモデルを洗練した．第三に，これらの技法を反復学習で統合し，各サイクルで異常スコアを再計算して擬似ラベルの精度を高め，検出精度を段階的に向上させた．

実験は未ラベル設定とラベルあり設定の双方で実施した．未ラベル設定では，本手法は機械タイプラベルのような粗い情報に依存する従来法を大きく上回った．ラベルあり設定では，選択的に取り込んだ外部データによって検出精度がさらに向上した．対象の正常音に対する類似度に基づいて選択した外部データを組み込むことは，常に検出精度の向上に寄与し，ランダム選択よりも優れていることを示した．関連する外部データが豊富な機械タイプ（$large\ N_{\mathrm{out}}\ machines$）では，外部データの最大数（$N_{\max}$）を学習データ規模に近い値（たとえば $N_{\max} = 1000$）に設定することが最適であり，過剰なデータ投入は性能劣化につながる．加えて，トリプレット損失を両段階で用いることで，stage 1 では擬似ラベルの品質が向上し，stage 2 ではラベル誤りの影響が緩和され，教師なし学習フレームワークにおける有効性が示された．

総じて，本手法はラベル利用が限定的な状況対して異常音検知性能を頑健に向上させ，産業応用において実践的かつ有効な解決策を提供することを確認した．

第5章

結論

本章では，第1章から第4章までの知見を実運用での判断材料として整理し，本研究の学術的貢献を総括する．最後に，残された課題と今後の展望を述べる．

5.1　運用時の設計指針

本節では，第2章から第4章で得られた知見を，実運用での意思決定に使いやすい形に整理する．現場で異常音検知システムを導入・運用する際には，以下の三つの典型的な判断が求められる：

1. **どのデータ収集が最も価値を持つか**：限られた予算と時間の中で，異常ラベル・属性ラベル・外部データのいずれを優先すべきか．
2. **モデルをどこで更新するか**：ドメインシフトや新規データ取得時に，前段（特徴量抽出器）と後段（異常検出器）のどちらを，どのタイミングで更新すべきか．
3. **どのように閾値を決めるか**：研究段階の評価指標（AUC / pAUC）と，運用段階の閾値設計（誤警報率の制御）をどう整合させるか．

表5.1は，データ可用性（異常ラベルyと属性ラベルcの有無）に基づく四象限A–Dそれぞれに対して，推奨される学習戦略と運用上の留意点をまとめたものである．

表 5.1: データ可用性に基づく実運用指針のまとめ．第 2–4 章の知見を統合し，各データ状況に応じた推奨学習戦略と運用上の留意点を示す．

区分	異常ラベル	属性ラベル	推奨される学習戦略	運用上の留意点
A	なし	あり	第 3 章：属性識別で埋め込みを学習し，後段で kNN や GMM によりスコア化する．擬似異常が用意できる場合は，OE を併用して前段を強化することもできる．	実異常がないため，B のように異常音を前段へ直接取り込む効果は使えない．ドメインシフトが支配的になりやすいので，まず後段更新（参照集合・統計量）を優先し，必要に応じて前段再学習を計画的に実施する．閾値は正常分布の分位点により FPR を管理する．
B	あり	あり	第 3 章：実異常と擬似異常を OE の負例として用い，さらに属性識別で正常側を細分化した埋め込みを得る．後段は属性ごとに GMM 等を構築する．	得られた異常音は前段学習へ反映する価値が高い．ただし停止時間・再配布制約がある場合は，再学習の実施頻度と反映ルール（いつ・どの程度の追加で更新するか）を事前に定める．
C	なし	なし	第 4 章：外部データから近傍サンプルを選別して導入し，擬似ラベル付与と反復再学習によりサブクラス構造を顕在化させる．その上で距離・尤度ベースの異常検出器を組み合わせる．	初期導入段階から適用できるが，外部データの混入リスクを前提に選別・反復の運用設計が必要である．反復再学習をいつ回すか（停止時間，計算場所，モデル配布）を運用フローに組み込む．
D	あり	なし	第 3 章：利用可能な異常音を OE の負例として前段学習に活用し，後段に距離・尤度ベース検知器を置く．属性ラベルがないため，必要に応じて擬似ラベルにより正常側の分割を補う．	属性情報がないと個体差や運転状態差を明示的に管理しにくい．可能なら運用の中で属性情報を付与する体制を整えるか，C 型（擬似ラベル／外部データ選別）を併用して正常側のサブクラス構造を補完する．

この表は，第 1 章で示した分類 (表 1.1) と，第 2–4 章で提案した手法を統合した実運用ガイドラインとして機能する．ここで Serial-OE は，実異常データを前段学習へ組み込める場合（B や D）に特に効果を発揮する．一方，A のように実異常が得られない場合でも，外部データや加工音により擬似異常が用意できるなら，OE を併用して埋め込みを強化する選択肢がある．

5.1.1 データ収集の優先順位

表 5.1 から読み取れる最も重要な示唆は，「少量でも異常データを記録できるなら，それは非常に高い価値を持つ」という点である．第 3 章の実験により，異常データが学習データの 0.1%（1 サンプル）でも有意な性能向上が得られることが示された．これは，C 領域（完全ラベルなし）でも外部データの選別と反復学習により異常音検知モデルとして成立するが，B 領域（異常データあり）に近づけられれば，費用対効果が格段に高まることを意味する．

したがって，システム導入時のデータ収集計画では，以下の優先順位で検討することが推奨される：

1. 稀少でも異常事象が発生した際の音声記録（B / D 領域への移行）
2. 個体差や運転条件を示す属性ラベルの付与（A / B 領域での性能向上）
3. 外部データの選別的導入（C 領域でのブートストラップ）

5.1.2 モデル更新のタイミングと方法

ドメインシフトが強い現場では，停止時間の制約が厳しい．第 2 章で整理したように，以下の段階的な更新戦略が現実的である：

1. **第一段階（軽量適応）**：後段の異常検出器（GMM / kNN）のみを更新，またはバッ

クエンド側の統計量（共分散，正規化パラメータ）のみを調整する．これにより，システム停止なしで運用条件の変化に追従できる．

2. **第二段階（前段再学習）**：定期メンテナンス時など，まとまった停止時間が確保できるタイミングで前段の特徴量抽出器を再学習する．第 3 章と第 4 章の枠組みは，この「いつ再学習を回すか」という運用設計を前提としている．

特に第 4 章で示した反復学習は，stage 2 から stage 3 への移行で平均 1 ポイント程度の性能向上が得られるため，少なくとも 3 回程度の反復を計画的に組み込むことが望ましい．

5.1.3　閾値設計と評価指標の整合

研究段階では AUC / pAUC による閾値非依存の性能評価が適しているが，運用段階では許容できる誤警報率（FPR）を定め，正常スコア分布の上側分位点から閾値を設計する．第 2 章で述べたように，この二本立てで評価することで，研究成果と運用目標を整合させ，現場での調整コストを抑えることができる．

具体的には，以下の手順を推奨する：

1. 開発段階：AUC / pAUC で手法間の系統比較を行う
2. 検証段階：目標 FPR を定め，正常データから閾値を算出する
3. 運用段階：実測 FPR と F1 スコアなどを併記し，必要に応じて閾値を微調整する

5.2　本研究の総括と意義

本節では，第 1 章から第 4 章までの内容を相互に関連付けて整理し，第 1.6 節で示した三つの貢献がどのように達成されたかを総括する．

5.2.1 貢献 1 の達成: 問題空間の体系的整理と設計原則の確立

第 1 章では，異常音検知を「環境軸」と「データ軸」という二つの観点から整理した．環境軸では，観測信号を

$$x(t) = h(t) * \{s_c(t) + n(t)\} \tag{5.1}$$

と定式化し，背景雑音 $n(t)$ や伝達系 $h(t)$ といった環境要因を不要因子として抑圧し，機械本体の挙動 $s_c(t)$ を安定して取り出すことが重要であるとした．これを運用上のドメインシフトの問題として捉え，以降の章の共通課題とした．

データ軸では，外部から与えられる情報源を異常ラベル y と属性ラベル c に集約し，それらが利用可能かどうかで学習条件を A–D の四象限（表 1.1）に分類した．A は半教師あり（正常データのみ，c あり），B は異常データありかつ c あり，C はラベルなし，D は異常データのみという設定である．この分類により，現場のデータ取得制約と研究設計を直接対応づけることが可能になった．

さらに，環境要因を抑圧しつつ機械挙動を強調する表現を前段で獲得し，その表現上で異常スコアを計算する後段を組み合わせる直列法を基本方針として定めた．この二段構成により，前段と後段を独立に更新・管理できる運用上の利点を確保し，第 2 章以降で展開するすべての手法の土台とした．

第 2 章では，この直列法を具体化し，前段の特徴量抽出器では属性識別や擬似異常との二値分類といった補助タスクを用いて環境要因の影響を抑え，後段では GMM や kNN といった距離・尤度ベースのスコアリングを行う実装を整理した．また，ドメインシフトへの対処として，現場データを用いた軽量な追従を行うドメイン適応と，追加適応なしでも壊れにくいモデルを目指すドメイン汎化を区別し，運用上の判断指針を示した．

これらの整理により，異常音検知における課題を「どの因子を抑圧し，どの因子を強調すべきか」という明確な問いに落とし込み，以降の章で提案する手法の共通言語を確立した．

5.2.2 貢献2の達成: データ制約下での具体的学習戦略の提案

第3章では，表1.1のB領域（異常データあり）およびD領域（異常データのみ）を対象とし，Serial-OEを提案した．Serial-OEは直列法を前提としつつ，前段の学習にOutlier Exposureに基づいた二値分類を組み込むことで，入手できた異常音や監視対象外の音を疑似異常として活用し，正常クラスタからの乖離を強調する埋め込みを獲得する．後段では，属性ラベルごとにGMMを当て，尤度に基づいて異常スコアを算出する．

実験により，異常データが学習データの0.1%（1サンプル）でも有意な性能向上が得られ，DCASE2020 Task 2データセットにおいて平均aAUC 93.54%を達成した．これは，従来の最良手法（Noisy-ArcMix, 91.98%）を1.56ポイント上回る結果である．この成果は，「異常データが少しでもあれば，それを後段だけでなく前段の埋め込み学習に反映させることで性能改善可能」という設計指針を実証するものである．

また，属性ラベルが欠落するDのような厳しい条件でも，異常データを疑似異常側に統合することで性能向上が見られ，この考え方が汎用的に有効であることを確認した．

第4章では，表1.1のC領域（完全ラベルなし）を対象とした．ここでは，直列法の前段を強化するために，

1. 外部データから対象機械に近いサンプルだけを選別して取り込む
2. データ間の関係から擬似ラベルを生成してサブクラス構造を浮かび上がらせる
3. これらを反復的に再学習する

という手続きを提示した．

DCASE2022–2024データセットを用いた実験により，未ラベル設定におけるベースライン（平均AUC 63.6%）を5.3ポイント上回る68.9%を達成し，ラベルあり設定（平均AUC 73.2%）との差を9.4ポイントから4.3ポイントまで縮小した．これにより，ラベルのない初期導入段階でも，環境要因よりも機械挙動そのものに基づいた埋め込み

表現を得て，後段の距離・尤度ベース検知を成立させられることを示した．

特に，トリプレット損失を併用することで擬似ラベルの品質が向上し，外部データの選別的導入により性能が安定することが確認された．反復学習では，stage 2 から stage 5 にかけて平均 AUC が 67.5%から 68.9%へと段階的に向上し，少なくとも 3 回程度の反復が有効であることが示された．

5.2.3 貢献 3 の達成: 実運用を見据えた設計指針の提示

表 5.1 では，上記の知見を実運用での意思決定に使いやすい形に整理した．各データ可用性条件（A–D）に応じて，どの種類の学習戦略が有効か，そして運用時にどこへ注意すべきかをまとめた．

この運用指針は，現場で直面する典型的な三つの意思決定に直結する:

- **どのデータ収集が最も価値を持つか**: C 領域の初期段階でも外部データの選別と反復学習により検知器として成立するが，B 領域に近づけられる（少量でも異常音を記録できる）なら，それは非常に高い価値を持つ．
- **モデルをどこで更新するか**: ドメインシフトが強く停止時間制約が厳しい現場では，まず後段の更新やバックエンド適応で追従し，余裕があるタイミングで前段の再学習を回す二段構えが現実的である．
- **どのように閾値を決めるか**: 研究段階の AUC に加えて，実運用では許容できる誤警報率を定め，正常スコア分布の上側分位点から閾値を引く二本立てで評価する．

さらに，第 2 章で整理したドメイン適応・汎化の観点と組み合わせることで，「前段・後段のどちらを，いつ，どのように更新するか」という具体的な運用フローを設計できるようになった．例えば，軽量な後段更新（統計量の再推定）は停止時間なしで実施可能であり，前段の再学習は定期メンテナンス時にまとめて行う，といった段階的適応戦略が実務的に有効であることを示した．

この設計指針は，単なるアルゴリズム比較にとどまらず，データ可用性と運用制約の両方に応じた判断材料を体系的に提供する点で，有用である．

5.2.4　本研究の意義

以上をまとめると，本研究は産業応用の異常音検知における課題を環境軸とデータ軸から体系的に分解し（貢献1），特にデータ軸では異常ラベルと属性ラベルの可用性に基づく四象限として整理した．その分類の中で現実的に重要だが従来整理の不十分だった領域（B, C, D）に対し，直列法という共通の設計原則のもとで具体的な学習戦略を提案した（貢献2）．さらに，これらの知見を実運用での判断材料として体系化した（貢献3）．

環境の揺らぎに起因するドメインシフト，ラベルの欠如，モデル更新や再配布の制約という産業的な現実を，直列法という共通の設計原則と四象限の分類を通じて結びつけたことが，本研究の最も本質的な貢献である．

5.3　残された課題と今後の展望

最後に，本研究で扱いきれなかった論点をまとめ，今後の展望を述べる．

5.3.1　ストリーミング運用と逐次的なモデル更新

第3章では，少量の異常データが得られた場合にそれを学習へ取り込む方法を示した．しかし，これは一定の単位で学習をやり直すことを前提としている．実際の設備監視では，システムを止めずに監視を続けながら，稀に得られる異常データや誤警報事例を人手で確認し，それを反映させたいという要求が強い．

第2章で述べた後段のみの更新，つまりスコア計算部や参照集合の更新による軽量

な適応はこの要求に沿いやすいが，前段の特徴量抽出器まで含めた再学習をオンラインで回す方法は十分に検討できていない．再学習を全停止メンテナンス時にまとめて行うのか，あるいはエッジ側で限定的に行うのかといった更新ポリシーと併せて検討する必要がある．

特に，継続学習の観点から，過去の知識を保持しつつ新規データに適応する手法 [140] を直列法の前段に組み込むことは，今後の重要な研究課題である．

5.3.2 エッジデバイスへの展開

本論文では性能そのものの議論を中心とし，推論時の計算資源やモデルサイズには限定的にしか触れていない．しかし実際には監視対象の機械ごとにセンサを取り付け，その場で異常スコアを計算するエッジ運用が求められる場合が多い．

直列法の長所は，前段と後段を分けられることにある．すなわち，前段の特徴量抽出器をある程度固定化してエッジ側に配布し，後段の軽量な検知器だけを現場ごとに更新するという分割が可能である．この構成を前提に，モデル圧縮 [141] や量子化 [141] などを組み合わせた具体的なデプロイ設計は，今後の課題として重要である．

また，エッジデバイス上でのリアルタイム推論を実現するためには，特徴量抽出の計算コストを削減する必要がある．例えば，第 4 章の複数解像度入力の一部を省略する，あるいは軽量なモバイルネット系アーキテクチャ [33] への置き換えなどが考えられる．

5.3.3 個別機械ごとの閾値と一元運用

本論文では，閾値は基本的に正常スコア分布の上側分位点で定めるという方針をとった．これは，現場の運用で「誤警報が多すぎると現場に受け入れられない」という要求に応えやすい．一方で，個体差が大きい機械では，ある機械だけ常にスコアが高く出るといったことが起こり得る．

その場合，単一閾値での一元運用と，個別閾値での機器別運用のどちらを選ぶべきかという実務上の判断が必要になる．本論文では，このトレードオフの定量化，および機器別閾値をどう維持するかという運用設計までは踏み込めていない．

一元運用の利点は管理コストの低さにあるが，個体差が大きい場合は機器別の動的閾値調整 [97] が有効である可能性がある．この点は，運用段階での継続的なモニタリングと併せて検討すべき課題である．

5.3.4　正常と異常が混在した未ラベル大規模ログ

本論文の第 4 章では，異常ラベルも属性ラベルも無いという厳しい条件 C を扱った．ただし，この設定では「学習用に与えられるデータは基本的に正常である」という前提を置いている．実際の現場では，長期運転ログをそのまま集めるだけで学習に使いたいという要求があり，このログには正常と異常が混ざっている場合がある．

その場合，異常を含んだサンプルまでを正常の一部として学習してしまうと，真に検知したい異常が「正常のバリエーション」として吸収されてしまう．これは本論文では扱っていない．この混在ログ条件は，教師なし異常検知における典型的な難題であり，本論文で示した C 領域の手法をそのままでは適用できない可能性がある．

混在状態から正常コアを自動的に抽出する処理，あるいは運転履歴や保全記録など非音響メタ情報を組み合わせたフィルタリング戦略が必要になると考えられる．例えば，Deep SVDD [44] や PaDiM [142] のような One-Class 学習手法を前段に組み込むことで，汚染に対する頑健性を高められる可能性がある．

第 3 章の実験 (図 3.5) では，正常データに異常データが混入した場合の性能劣化を評価したが，これは「混入した異常データを事後的に検出・除去する」手法ではなく，「混入に対してどこまで頑健か」を測る実験であった．今後は，混入異常の自動検出や，汚染に頑健な学習アルゴリズム [143] の開発が求められる．

5.3.5 性能上限と汎用性

本論文では，精度 100 パーセントという意味での完全な検知は当然ながら達成していない．また，ドメインシフトを完全に抑え込み，かつ停止時間ゼロで常時アップデートできるような万能な運用設計も提示できていない．

しかし，本論文は環境軸とデータ軸の二つの視点から，現実的に遭遇しうるデータ可用性の領域を A–D に分け，それぞれの条件に対してどのように直列法を設計・拡張すべきかを具体的に示した．この点は，今後の産業応用に向けての足場として機能すると考える．

特に，第 5.1 節で示した運用指針は，「どのようなデータがどこまで手に入るか」という制約に向き合い，その制約下で異常音検知システムをどう設計して運用に載せるかを体系立てて整理したものである．この枠組みは，機械音だけでなく，振動データや画像データなど他のモダリティにも応用可能な汎用性を持つと考えられる．

以上を総合すると，本論文は「どのようなデータがどこまで手に入るか」という制約に向き合い，その制約下で異常音検知システムをどう設計して運用に載せるかを体系立てて示した．環境の揺らぎに起因するドメインシフト，ラベルの欠如，モデル更新や再配布の制約という産業的な現実を，直列法という共通の設計原則と四象限の分類を通じて結びつけたことが，本研究の最も本質的な貢献である．

付録

ドメインシフト下での詳細性能評価

本付録では，第 4 章で提案した各手法について，ドメインシフト環境下での検出性能を補足的に示す．ここでいうドメインシフトとは，同一の機械タイプに対して収録条件や設置環境などが異なる ソースドメインとターゲットドメインの差異を指す．本論文の主眼は，属性ラベルを用いない設定において異常検知性能を改善できるかどうかであり，第 4 章では主にその観点で議論を行った．一方，実運用ではソースとは異なるターゲット環境での推論が避けられないため，ソース／ターゲットそれぞれでの AUC を併記し，提案手法がドメイン間でどの程度性能を維持できるかを評価することも重要となる．表 5.2, 5.3 では，DCASE 2022〜2024 Task 2 データセットを用い，属性ラベルを利用しない条件（w/o label）および属性ラベルを利用可能な条件（w/ label）における平均 AUC [%] を ソース／ターゲット ドメインごとに報告する．各値は 5 つのランダムシードに対する平均 ± 標準偏差である．

表 5.2: DCASE 2022〜2024 Task 2 データセットにおける，各手法のラベルなし（属性非利用）条件下での AUC [%] の平均値である．ここで“ソース”と“ターゲット”は 2 つのドメインを示し，値は 5 つのランダムシードに対する平均値 ± 標準偏差である．

DCASE	Use label	Method	stage	development		evaluation	
				ソース	ターゲット	ソース	ターゲット
2022	w/ label	Wilkinghoff [125]	1	**86.0±0.9**	78.2±0.7	77.7±0.8	71.6±1.0
		Ba [3]	1	84.9±0.6	78.6±1.7	**80.2±0.6**	**74.2±1.0**
	w/o label	Wilkinghoff [125]	1	69.6±6.1	64.2±5.3	66.9±5.5	63.0±2.2
		Ba [3]	1	71.5±1.2	71.1±1.2	70.4±1.6	66.2±1.4
		Ba+$\mathcal{L}_{\mathrm{trp}}$	1	72.1±1.8	71.7±1.4	70.5±1.6	67.1±0.9
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ex	2	73.6±1.0	74.9±0.8	71.8±0.5	67.4±1.1
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps	2	79.5±1.0	**75.1±0.9**	75.5±0.6	**69.6±1.0**
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	2	79.0±1.1	73.3±1.3	76.4±1.4	68.8±1.2
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	3	80.8±2.0	72.7±2.1	**76.7±0.8**	68.5±2.3
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	4	80.2±2.2	72.8±1.8	76.5±0.4	69.2±0.8
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	5	**82.9±0.6**	74.5±1.8	76.3±0.7	68.9±0.6
2023	w/ label	Wilkinghoff [125]	1	71.2±1.6	75.0±1.5	75.5±0.8	68.7±2.2
		Ba [3]	1	72.0±1.4	74.7±1.5	78.0±1.5	68.3±2.1
	w/o label	Wilkinghoff [125]	1	64.9±1.8	63.6±0.7	62.8±0.8	56.7±1.7
		Ba [3]	1	65.7±1.5	63.5±1.2	60.6±0.9	57.3±1.7
		Ba+$\mathcal{L}_{\mathrm{trp}}$	1	64.6±2.5	64.8±1.4	60.8±1.0	57.8±2.5
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ex	2	65.7±1.7	64.2±1.1	62.0±1.5	58.9±2.2
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps	2	69.5±2.0	67.5±1.2	63.1±1.9	67.9±2.3
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	2	70.0±1.3	67.5±2.5	65.8±1.5	66.2±4.2
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	3	71.1±1.5	**69.3±2.3**	65.9±0.5	70.3±1.7
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	4	71.8±1.2	68.3±1.8	68.6±1.6	**73.3±1.0**
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	5	**72.0±1.7**	68.3±1.1	**68.9±0.8**	72.7±2.0
2024	w/ label	Wilkinghoff [125]	1	68.9±1.3	63.8±1.8	63.2±1.4	62.7±1.7
		Ba [3]	1	74.6±1.1	64.5±1.8	64.0±1.4	65.3±2.4
	w/o label	Wilkinghoff [125]	1	65.9±1.3	58.8±1.4	54.2±1.4	57.4±0.9
		Ba [3]	1	66.1±3.2	59.5±1.8	52.5±1.5	54.6±0.8
		Ba+$\mathcal{L}_{\mathrm{trp}}$	1	65.4±1.9	59.0±1.0	52.6±2.1	55.6±1.6
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ex	2	68.2±2.4	**61.4±0.9**	54.7±1.1	55.2±0.8
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps	2	68.8±1.8	59.6±0.9	57.0±1.8	56.5±1.3
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	2	68.2±1.4	59.8±1.4	58.6±1.1	55.6±0.9
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	3	**71.6±0.0**	58.2±1.1	58.5±1.0	56.2±0.1
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	4	69.9±1.0	59.2±2.7	55.7±0.9	**57.5±0.7**
		Ba+$\mathcal{L}_{\mathrm{trp}}$+Ps+Ex	5	69.9±2.2	59.9±0.4	**59.1±1.4**	56.9±0.7

表 5.3: DCASE 2022〜2024 Task 2 データセットにおける，各手法のラベルあり（属性利用可能）条件下での AUC [%] の平均値である．表記は 表 5.2 と同一である．

DCASE	Method	stage	development		evaluation	
			ソース	ターゲット	ソース	ターゲット
2022	Wilkinghoff [125]	1	**86.0±0.9**	78.2±0.7	77.7±0.8	71.6±1.0
	Ba [3]	1	84.9±0.6	78.6±1.7	**80.2±0.6**	**74.2±1.0**
	Ba+$\mathcal{L}_{\mathrm{trp}}$	1	83.7±0.6	76.4±1.1	71.2±0.5	66.2±0.5
	Ba+Ex	2	84.7±0.4	**79.4±0.7**	79.9±0.8	73.3±0.8
	Ba+Ps	2	82.9±0.3	76.4±1.4	79.8±0.1	72.4±1.0
	Ba+Ps+$\mathcal{L}_{\mathrm{trp}}$	2	82.7±0.7	77.6±0.8	79.1±0.5	72.4±0.5
	Ba+Ex	3	84.9±0.9	79.0±1.1	79.5±0.7	73.9±0.5
2023	Wilkinghoff [125]	1	71.2±1.6	75.0±1.5	75.5±0.8	68.7±2.2
	Ba [3]	1	72.0±1.4	74.7±1.5	78.0±1.5	68.3±2.1
	Ba+$\mathcal{L}_{\mathrm{trp}}$	1	70.5±2.8	74.3±2.4	71.5±1.7	66.4±1.2
	Ba+Ex	2	**72.1±1.3**	**77.2±0.8**	**79.0±0.3**	**69.2±1.7**
	Ba+Ps	2	71.3±1.0	**77.2±1.4**	75.5±2.4	67.3±2.1
	Ba+Ps+$\mathcal{L}_{\mathrm{trp}}$	2	69.4±0.7	74.9±1.2	70.9±1.9	67.3±2.5
	Ba+Ex	3	70.9±0.8	76.6±1.4	**79.0±1.1**	69.0±1.6
2024	Wilkinghoff [125]	1	68.9±1.3	63.8±1.8	63.2±1.4	62.7±1.7
	Ba [3]	1	74.6±1.1	64.5±1.8	64.0±1.4	65.3±2.4
	Ba+$\mathcal{L}_{\mathrm{trp}}$	1	70.9±1.4	65.4±0.7	60.9±2.9	64.7±1.9
	Ba+Ex	2	**75.2±0.2**	**65.5±1.5**	**64.6±2.2**	**67.3±1.0**
	Ba+Ps	2	72.5±0.7	**65.5±1.9**	60.9±1.1	63.0±2.0
	Ba+Ps+$\mathcal{L}_{\mathrm{trp}}$	2	71.3±1.9	64.9±1.7	59.6±1.5	64.7±1.0
	Ba+Ex	3	74.2±1.5	64.7±1.2	64.5±2.2	67.1±0.8

謝辞

本博士論文は，名古屋大学大学院情報学研究科に在籍した期間を通じて行ってきた研究の成果をまとめたものであり，これまでの研究生活を支えてくださった多くの方々のご指導とご支援なくして完成することはできませんでした．ここに，お世話になった皆様に深く感謝の意を表します．

まず，本博士論文の審査をお引き受けくださった，名古屋大学 未来社会創造機構 武田一哉教授，同大学 情報基盤センター 戸田智基教授，京都大学 大学院情報学研究科 井本桂右准教授の 3 名の先生方に心より御礼申し上げます．ご多忙の中，ご査読と貴重なご助言を賜り，本論文を学術的により価値あるものへと導いていただきました．

指導教員である武田一哉教授には，日頃より研究に専念できる自由で恵まれた研究環境を提供していただきました．直接ご指導いただく機会は多くはなかったものの，武田研究室という大きな枠組みとネットワークの中で，多様な研究者と議論し視野を広げる貴重な機会に恵まれました．このような環境のもとで博士課程の研究に取り組むことができたことに，深く感謝いたします．

学部生の頃より一貫して丁寧なご指導を賜りました戸田智基教授には，研究テーマの設定から論文の書き方，発表の仕方に至るまで，研究者としての基礎を一つ一つ教えていただきました．研究が停滞した際にも粘り強く議論に付き合ってくださり，進むべき方向を示していただいたおかげで，今日まで研究を続けることができました．長年にわたるご指導に，心より感謝申し上げます．

本博士論文の内容について大変丁寧に査読してくださった井本桂右准教授に厚く御

礼申し上げます．いただいた多くの具体的なご指摘やご助言は，本論文の構成や議論の精緻化において非常に有益であり，それらを通じて自らの研究を客観的に見つめ直す貴重な機会を得ることができました．

また，ゼミにおいて，貴重なご意見とご助言をくださった名古屋大学 情報学研究科 藤井慶輔准教授，東京大学 大学院情報学環 石黒祥生准教授，名古屋大学 情報学研究科 特任助教であり，株式会社 Human Dataware Lab. 代表取締役 CEO でもある大谷健登特任助教，名古屋大学 大学院情報学研究科 HUANG Wen-Chin 助教にも深く感謝いたします．先生方との議論を通じて多くの刺激をいただき，新たな視点から自分の研究を見つめ直す機会となりました．

さらに，株式会社 Human Dataware Lab. 取締役 COO 林智樹さんには，Human Dataware Lab. における実務面でのご指導に加え，研究室においても私の研究に深く関わっていただきました．研究の細かな部分にまで目を配って一緒に考えてくださり，行き詰まった際には具体的な助言をいただきました．林さんの支えがなければ，ここまで研究を進めることはできませんでした．ここに改めて深く感謝申し上げます．

さらに，名古屋大学 大学院情報学研究科 博士課程の後輩である藤村拓弥さんには，博士課程 3 年間を通じて DCASE Task 2 Challenge に共著者としてともに挑戦してくれました．困難な課題に取り組む過程で，日々の議論や実験を通じて多くの刺激を与えてもらい，大きな励みとなりました．藤村さんとの共同研究は，本博士論文の成果の一部を形作る重要な要素であり，ここに感謝の意を表します．

株式会社 KPMG Forensic & Risk Advisory の皆様には，博士課程への進学と研究と業務の両立に対して深いご理解を示していただきました．繁忙な業務の中でも，挑戦の機会を与えていただいたおかげで，本論文の執筆と研究を継続することができました．社会人としての成長と研究者としての挑戦を両立する貴重な環境を提供してくださったことに，心より感謝いたします．

あわせて，これまで在籍した研究室の先輩方，同期の仲間，そして後輩の皆さんに

も深く感謝いたします．日々の議論や雑談を通じて，多くの刺激と学びを得ることができました．研究室というコミュニティでともに過ごした時間は，私の博士課程生活を非常に豊かなものにしてくれました．

長い学生生活を送り，博士課程まで進学することを温かく見守り，精神面・生活面の両方から支えてくれた家族に，心から感謝いたします．どのような状況でも私を信じて応援し続けてくれたことが，大きな支えとなりました．

そして何よりも，私を博士課程に誘い，ともに苦楽を分かち合い，同級生として一緒に学び，さらに伴侶として常にそばで支え続けてくれた妻・花奈には，深甚なる感謝の意を表します．研究が思うように進まず落ち込んだときも，ともに励まし合い，前を向く力を与えてくれました．花奈の存在があったからこそ，博士課程を最後まで走り抜け，本論文を完成させることができました．

以上，ここに記した全ての方々に，改めて深く感謝申し上げます．

参考文献

[1] K. Wilkinghoff, “Sub-Cluster AdaCos: Learning Representations for Anomalous Sound Detection,” in *Proc. International Joint Conference on Neural Networks*, 2021, pp. 1–8.

[2] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” in *Proc. Computer Vision and Pattern Recognition*, 2019, pp. 4685–4694.

[3] T. Fujimura, I. Kuroyanagi, and T. Toda, “Improvements of Discriminative Feature Space Training for Anomalous Sound Detection in Unlabeled Conditions,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2025, pp. 1–5.

[4] Y. Koizumi, S. Saito, H. Uematsu, and N. Harada, “Optimizing acoustic feature extractor for anomalous sound detection based on Neyman-Pearson lemma,” in *Proc. European Signal Processing Conference*, 2017, pp. 698–702.

[5] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi, and N. Harada, “Unsupervised Detection of Anomalous Sound Based on Deep Learning and the Neyman–Pearson Lemma,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, pp. 212–224, 2019.

[6] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin, "Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges," *IEEE Access*, vol. 6, pp. 6505–6519, 2018.

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009, 58 pages.

[8] C. C. Aggarwal, *Outlier Analysis*, 2nd ed. 2017.

[9] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and Discussion on DCASE 2021 Challenge Task 2: Unsupervised Anomalous Detection for Machine Condition Monitoring Under Domain Shifted Conditions," in *Proc. Detection and Classification of Acoustic Scenes and Events 2021 Workshop*, 2021, pp. 186–190.

[10] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, and Y. Kawaguchi, "Description and Discussion on DCASE 2022 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques," in *Proc. Detection and Classification of Acoustic Scenes and Events 2022 Workshop*, 5 pages, 2022.

[11] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Description and Discussion on DCASE 2023 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring," in *Proc. Detection and Classification of Acoustic Scenes and Events 2023 Workshop*, 2023, pp. 31–35.

[12] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Descrip-

tion and Discussion on DCASE 2024 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring," in *Proc. Detection and Classification of Acoustic Scenes and Events 2024 Workshop*, 2024, pp. 111–115.

[13] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. Harada, "Description and Discussion on DCASE2020 Challenge Task2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring," in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 81–85.

[14] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Anomalous Sound Detection Based on Interpolation Deep Neural Network," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 271–275.

[15] R. Giri, F. Cheng, K. Helwani, S. V. Tenneti, U. Isik, and A. Krishnaswamy, "Group Masked Autoencoder Based Density Estimator for Audio Anomaly Detection," in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 51–55.

[16] H. Purohit, R. Tanabe, T. Endo, K. Suefusa, Y. Nikaido, and Y. Kawaguchi, "Deep Autoencoding GMM-Based Unsupervised Anomaly Detection in Acoustic Signals and its Hyper-Parameter Optimization," in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 175–179.

[17] S. Kapka, "ID-Conditioned Auto-Encoder for Unsupervised Anomaly Detection," in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 71–75.

[18] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, and G. L. Foresti, "VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization," in *Proc. International Symposium on Industrial Electronics*, 2021, pp. 1–6.

[19] R. Müller, S. Illium, and C. Linnhoff-Popien, "Deep Recurrent Interpolation Networks for Anomalous Sound Detection," in *Proc. International Joint Conference on Neural Networks*, 2021, pp. 1–7.

[20] T. Hayashi, T. Komatsu, R. Kondo, T. Toda, and K. Takeda, "Anomalous sound event detection based on WavNnet," in *Proc. European Signal Processing Conference*, 2018, pp. 2494–2498.

[21] X. Xia, X. Pan, N. Li, X. He, L. Ma, X. Zhang, and N. Ding, "GAN-Based Anomaly Detection: A Review," *Neurocomputing*, vol. 493, no. C, pp. 497–535, 2022.

[22] A. Jiang, W.-Q. Zhang, Y. Deng, P. Fan, and J. Liu, "Unsupervised Anomaly Detection and Localization of Machine Audio: A Gan-Based Approach," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.

[23] D. Reynolds, "Gaussian Mixture Models," in *Encyclopedia of Biometrics*. 2009, pp. 659–663.

[24] W. Liu, D. Cui, Z. Peng, and J. Zhong, "Outlier Detection Algorithm Based on Gaussian Mixture Model," in *Proc. International Conference on Power, Intelligent Computing and System*, 2019, pp. 488–492.

[25] J. Guan, Y. Liu, Q. Zhu, T. Zheng, J. Han, and W. Wang, "Time-Weighted Frequency Domain Audio Representation with GMM Estimator for Anomalous

Sound Detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.

[26] G. Papamakarios, T. Pavlakou, and I. Murray, "Masked Autoregressive Flow for Density Estimation," in *Proc. International Conference on Neural Information Processing Systems*, 2017, pp. 2335–2344.

[27] P. Kirichenko, P. Izmailov, and A. G. Wilson, "Why Normalizing Flows Fail to Detect Out-of-Distribution Data," in *Proc. International Conference on Neural Information Processing Systems*, 2020, pp. 20 578–20 589.

[28] K. Dohi, T. Endo, H. Purohit, R. Tanabe, and Y. Kawaguchi, "Flow-Based Self-Supervised Density Estimation for Anomalous Sound Detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 336–340.

[29] D. Kim, S. Baik, and T. H. Kim, "Sanflow: Semantic-aware normalizing flow for anomaly detection," in *Proc. Neural Information Processing Systems*, vol. 36, 2023, pp. 75 434–75 454.

[30] M. Rudolph, B. Wandt, and B. Rosenhahn, " Same Same But DifferNet: Semi-Supervised Defect Detection with Normalizing Flows," in *Proc. Winter Conference on Applications of Computer Vision*, 2021, pp. 1906–1915.

[31] T. Inoue, P. Vinayavekhin, S. Morikuni, S. Wang, T. Hoang Trong, D. Wood, M. Tatsubori, and R. Tachibana, "Detection of Anomalous Sounds for Machine Condition Monitoring using Classification Confidence," in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 66–70.

[32] R. Giri, S. V. Tenneti, F. Cheng, K. Helwani, U. Isik, and A. Krishnaswamy, "Self-Supervised Classification for Detecting Anomalous Sounds," in *Proc. De-*

tection and Classification of Acoustic Scenes and Events 2020 Workshop, 2020, pp. 46–50.

[33] Y. Liu, J. Guan, Q. Zhu, and W. Wang, "Anomalous Sound Detection Using Spectral-Temporal Information Fusion," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 816–820.

[34] K. Wilkinghoff, "Combining Multiple Distributions based on Sub-Cluster AdaCos for Anomalous Sound Detection under Domain Shifted Conditions," in *Proc. Detection and Classification of Acoustic Scenes and Events 2021 Workshop*, 2021, pp. 55–59.

[35] K. Wilkinghoff, "Design Choices for Learning Embeddings from Auxiliary Tasks for Domain Generalization in Anomalous Sound Detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.

[36] H. Chen, L. Ran, X. Sun, and C. Cai, "SW-WAVENET: Learning Representation from Spectrogram and Wavegram Using Wavenet for Anomalous Sound Detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.

[37] K. Wilkinghoff and F. Kurth, "Why Do Angular Margin Losses Work Well for Semi-Supervised Anomalous Sound Detection?" *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 608–622, 2024.

[38] H. Hojjati and N. Armanfard, "Self-Supervised Acoustic Anomaly Detection Via Contrastive Learning," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 3253–3257.

[39] X.-M. Zeng, Y. Song, Z. Zhuo, Y. Zhou, Y.-H. Li, H. Xue, L.-R. Dai, and I. McLoughlin, "Joint generative-contrastive representation learning for anoma-

lous sound detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.

[40] J. Guan, F. Xiao, Y. Liu, Q. Zhu, and W. Wang, "Anomalous Sound Detection Using Audio Representation with Machine ID Based Contrastive Learning Pretraining," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.

[41] S. Choi and J.-W. Choi, "Noisy-Arcmix: Additive Noisy Angular Margin Loss Combined With Mixup For Anomalous Sound Detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 516–520.

[42] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Anomalous Sound Detection Using a Binary Classification Model and Class Centroids," in *Proc. European Signal Processing Conference*, 2021, pp. 1995–1999.

[43] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support Vector Method for Novelty Detection," in *Proc. International Conference on Neural Information Processing Systems*, 1999, pp. 582–588.

[44] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep One-Class Classification," in *Proc. International Conference on Machine Learning*, 2018, pp. 4393–4402.

[45] L. Ruff, R. A. Vandermeulen, N. Görnitz, A. Binder, E. Müller, K.-R. Müller, and M. Kloft, "Deep Semi-Supervised Anomaly Detection," in *Proc. International Conference on Learning Representations*, 23 pages, 2020.

[46] P. Primus, V. Haunschmid, P. Praher, and G. Widmer, "Anomalous Sound Detection as a Simple Binary Classification Problem with Careful Selection

of Proxy Outlier Examples," in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 170–174.

[47] I. Kuroyanagi, T. Hayashi, Y. Adachi, T. Yoshimura, K. Takeda, and T. Toda, "An Ensemble Approach to Anomalous Sound Detection Based on Conformer-Based Autoencoder and Binary Classifier Incorporated with Metric Learning," in *Proc. Detection and Classification of Acoustic Scenes and Events 2021 Workshop*, 2021, pp. 110–114.

[48] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Improvement of Serial Approach to Anomalous Sound Detection by Incorporating Two Binary Cross-Entropies for Outlier Exposure," in *Proc. European Signal Processing Conference*, 2022, pp. 294–298.

[49] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. International Conference on Machine Learning*, 2017, pp. 1321–1330.

[50] K. Wilkinghoff and A. Cornaggia-Urrigshardt, "On Choosing Decision Thresholds for Anomalous Sound Detection in Machine Condition Monitoring," in *Proc. International Congress on Acoustics*, 12 pages, 2022.

[51] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. S. Yu, "Generalizing to Unseen Domains: A Survey on Domain Generalization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 8052–8072, 2023.

[52] B. Sun and K. Saenko, "Deep CORAL: Correlation Alignment for Deep Domain Adaptation," in *European Conference on Computer Vision*, 2016, pp. 443–450.

[53] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[54] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient Algorithms for Mining Outliers from Large Data Sets," *SIGMOD Rec.*, vol. 29, no. 2, pp. 427–438, 2000.

[55] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," *SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, 2000.

[56] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proc. Computer Vision and Pattern Recognition*, 2022, pp. 14 298–14 308.

[57] Y. Wang, N. J. Bryan, M. Cartwright, J. Pablo Bello, and J. Salamon, "Few-shot continual learning for audio classification," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 321–325.

[58] Y. Zeng, H. Liu, L. Xu, Y. Zhou, and L. Gan, "ROBUST ANOMALY SOUND DETECTION FRAMEWORK FOR MACHINE CONDITION MONITORING," DCASE2022 Challenge, Tech. Rep., 2022, 3 pages.

[59] K. T. Mai, T. Davies, L. D. Griffin, and E. Benetos, "Explaining the Decision of Anomalous Sound Detectors," in *Proc. Detection and Classification of Acoustic Scenes and Events 2022 Workshop*, 5 pages, 2022.

[60] J. Park and S. Yoo, "DCASE 2020 Task2: Anomalous Sound Detection using Relevant Spectral Feature and Focusing Techniques in the Unsupervised Learning Scenario," in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 140–144.

[61] K. Shimonishi, K. Dohi, and Y. Kawaguchi, “Anomalous Sound Detection Based on Sound Separation,” in *Proc. Interspeech*, 2023, pp. 2733–2737.

[62] S. Perez-Castanos, J. Naranjo-Alcazar, P. Zuccarello, and M. Cobos, “Anomalous Sound Detection using Unsupervised and Semi-Supervised Autoencoders and Gammatone Audio Representation,” in *Proc. Detection and Classification of Acoustic Scenes and Events 2020 Workshop*, 2020, pp. 145–149.

[63] J. A. Lopez, G. Stemmer, P. Lopez Meyer, P. Singh, J. Del Hoyo Ontiveros, and H. Cordourier, “Ensemble of Complementary Anomaly Detectors Under Domain Shifted Conditions,” in *Proc. Detection and Classification of Acoustic Scenes and Events 2021 Workshop*, 2021, pp. 11–15.

[64] X. Cai and H. Dinkel, “A Contrastive Semi-Supervised Learning Framework For Anomaly Sound Detection,” in *Proc. Detection and Classification of Acoustic Scenes and Events 2021 Workshop*, 2021, pp. 31–34.

[65] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proc. International Conference on Machine Learning*, 2015, pp. 448–456.

[66] X. Zhang, R. Zhao, Y. Qiao, X. Wang, and H. Li, “AdaCos: Adaptively Scaling Cosine Logits for Effectively Learning Deep Face Representations,” in *Proc. Computer Vision and Pattern Recognition*, 2019, pp. 10 815–10 824.

[67] K. Wilkinghoff, “AdaProj: Adaptively Scaled Angular Margin Subspace Projections for Anomalous Sound Detection with Auxiliary Classification Tasks,” in *Proc. Detection and Classification of Acoustic Scenes and Events 2024 Workshop*, 2024, pp. 186–190.

[68] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “Mixup: Beyond Empirical Risk Minimization,” in *Proc. International Conference on Learning Representations*, 13 pages, 2018.

[69] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, “Improvement of anomalous sound detection method considering the distribution of embedding,” in *Proc. International Congress on Acoustics*, 5 pages, 2022.

[70] Y. Tachioka, “Outlier Exposure with Efficient Division of Positive and Negative Examples for Anomalous Sound Detection,” in *Proc. European Signal Processing Conference*, 2024, pp. 76–80.

[71] Y. Zhang, S. Hongbin, Y. Wan, and M. Li, “Outlier-aware inlier modeling and multi-scale scoring for anomalous sound detection via multitask learning,” in *Proc. Interspeech*, 2023, pp. 5381–5385.

[72] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, “Panns: Large-scale pretrained audio neural networks for audio pattern recognition,” *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 28, pp. 2880–2894, 2020.

[73] S. Chen, Y. Wu, C. Wang, S. Liu, D. Tompkins, Z. Chen, W. Che, X. Yu, and F. Wei, “BEATs: Audio Pre-Training with Acoustic Tokenizers,” in *Proc. International Conference on Machine Learning*, vol. 202, 2023, pp. 5178–5193.

[74] W. Chen, Y. Liang, Z. Ma, Z. Zheng, and X. Chen, “EAT: Self-Supervised Pre-Training with Efficient Audio Transformer,” in *Proc. International Joint Conference on Artificial Intelligence*, Main Track, 2024, pp. 3807–3815.

[75] A. L. Cramer, H.-H. Wu, J. Salamon, and J. P. Bello, “Look, Listen, and Learn More: Design Choices for Deep Audio Embeddings,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 3852–3856.

[76] E. J. Hu, yelong shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “LoRA: Low-Rank Adaptation of Large Language Models,” in *Proc. International Conference on Learning Representations*, 13 pages, 2022.

[77] X. Zheng, A. Jiang, B. Han, Y. Qian, P. Fan, J. Liu, and W.-Q. Zhang, “Improving Anomalous Sound Detection Via Low-Rank Adaptation Fine-Tuning of Pre-Trained Audio Models,” in *Proc. Spoken Language Technology*, 2024, pp. 969–974.

[78] A. Jiang, B. Han, Z. Lv, Y. Deng, W.-Q. Zhang, X. Chen, Y. Qian, J. Liu, and P. Fan, “AnoPatch: Towards Better Consistency in Machine Anomalous Sound Detection,” in *Proc. Interspeech*, 2024, pp. 107–111.

[79] K. Dohi and Y. Kawaguchi, “Distributed Collaborative Anomalous Sound Detection by Embedding Sharing,” in *Proc. European Signal Processing Conference*, 2024, pp. 91–95.

[80] X. Zheng, A. Jiang, B. Han, Y. Qian, P. Fan, J. Liu, and W. Zhang, “Improving Anomalous Sound Detection Via Low-Rank Adaptation Fine-Tuning of Pre-Trained Audio Models,” in *Proc. Spoken Language Technology*, 2024, pp. 969–974.

[81] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazare, M. Lomeli, L. Hosseini, and H. Jégou, “The Faiss Library,” *IEEE Transactions on Big Data*, pp. 1–17, 2025.

[82] J. Johnson, M. Douze, and H. Jégou, “Billion-scale similarity search with GPUs,” *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547, 2019.

[83] P. J. Rousseeuw and B. C. van Zomeren, “Unmasking Multivariate Outliers and Leverage Points,” *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 633–639, 1990.

[84] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, “First-Shot Anomaly Sound Detection for Machine Condition Monitoring: A Domain Generalization Baseline,” in *Proc. European Signal Processing Conference*, 2023, pp. 191–195.

[85] F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, and S. Rota Bulò, “AutoDIAL: Automatic DomaIn Alignment Layers,” in *Proc. International Conference on Computer Vision*, 2017, pp. 5077–5085.

[86] H. Chen, Y. Song, L.-R. Dai, I. McLoughlin, and L. Liu, “Self-Supervised Representation Learning for Unsupervised Anomalous Sound Detection Under Domain Shift,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 471–475.

[87] B. Chen, L. Bondi, and S. Das, “Learning to adapt to domain shifts with few-shot samples in anomalous sound detection,” in *Proc. International Conference on Pattern Recognition*, 2022, pp. 133–139.

[88] K. Wilkinghoff, T. Fujimura, K. Imoto, and J. Le Roux, “Handling Domain Shifts for Anomalous Sound Detection: A Review,” in *Proc. 51st Annual Meeting on Acoustics*, 2025, pp. 101–104.

[89] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: Synthetic minority over-sampling technique,” *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[90] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, “Two-stage anomalous sound detection systems using domain generalization and specialization techniques,” DCASE2022 Challenge, Tech. Rep., 2022, 5 pages.

[91] I. Nejjar, J. Meunier-Pion, G. Frusque, and O. Fink, “DG-Mix: Domain Generalization for Anomalous Sound Detection Based on Self-Supervised Learning,” in *Proc. Detection and Classification of Acoustic Scenes and Events 2022 Workshop*, 2022, 5 pages.

[92] J. Yan, Y. Cheng, Q. Wang, L. Liu, W. Zhang, and B. Jin, “Transformer and Graph Convolution-Based Unsupervised Detection of Machine Anomalous Sound Under Domain Shifts,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 4, pp. 2827–2842, 2024.

[93] S. Venkatesh, G. Wichern, A. Subramanian, and J. Le Roux, “Improved Domain Generalization via Disentangled Multi-Task Learning in Unsupervised Anomalous Sound Detection,” in *Proc. Detection and Classification of Acoustic Scenes and Events 2022 Workshop*, 2022, 5 pages.

[94] J. Guan, J. Tian, Q. Zhu, F. Xiao, H. Zhang, and X. Liu, “Disentangling Hierarchical Features for Anomalous Sound Detection Under Domain Shift,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2025, pp. 1–5.

[95] K. Dohi, T. Endo, and Y. Kawaguchi, “Disentangling physical parameters for anomalous sound detection under domain shifts,” in *Proc. European Signal Processing Conference*, 2022, pp. 279–283.

[96] H. Lan, Q. Zhu, J. Guan, Y. Wei, and W. Wang, “Hierarchical metadata information constrained self-supervised learning for anomalous sound detection

under domain shift," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 7670–7674.

[97] K. Wilkinghoff, H. Yang, J. Ebbers, F. G. Germain, G. Wichern, and J. L. Roux, "Keeping the balance: Anomaly score calculation for domain generalization," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2025, pp. 1–5.

[98] P. Saengthong and T. Shinozaki, "Deep generic representations for domain-generalized anomalous sound detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2025, pp. 1–5.

[99] K. Wilkinghoff and K. Imoto, "F1-ev score: Measuring the likelihood of estimating a good decision threshold for semi-supervised anomaly detection," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 256–260.

[100] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Serial-OE: Anomalous Sound Detection Based on Serial Method with Outlier Exposure Capable of Using Small Amounts of Anomalous Data for Training," *APSIPA Transactions on Signal and Information Processing*, vol. 14, no. 1, 32 pages, 2025.

[101] T. V. Ho, K. Dohi, and Y. Kawaguchi, "Stream-based Active Learning for Anomalous Sound Detection in Machine Condition Monitoring," in *Proc. Interspeech*, 2024, pp. 102–106.

[102] Y. Koizumi, S. Murata, N. Harada, S. Saito, and H. Uematsu, "SNIPER: Few-shot Learning for Anomaly Detection to Minimize False-negative Rate with Ensured True-positive Rate," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 915–919.

[103] Y. Koizumi, M. Yasuda, S. Murata, S. Saito, H. Uematsu, and N. Harada, "SPIDERnet: Attention Network For One-Shot Anomaly Detection In Sounds," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2020, pp. 281–285.

[104] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller, "A Unifying Review of Deep and Shallow Anomaly Detection," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 756–795, 2021.

[105] K. Sohn, C.-L. Li, J. Yoon, M. Jin, and T. Pfister, "Learning and Evaluating Representations for Deep One-Class Classification," in *Proc. International Conference on Learning Representations*, 32 pages, 2021.

[106] D. Hendrycks, M. Mazeika, and T. Dietterich, "Deep Anomaly Detection with Outlier Exposure," in *Proc. International Conference on Learning Representations*, 18 pages, 2019.

[107] C. Ding, G. Pang, and C. Shen, "Catching Both Gray and Black Swans: Open-set Supervised Anomaly Detection," in *Proc. Computer Vision and Pattern Recognition*, 2022, pp. 7378–7388.

[108] S. Yun, D. Han, S. Chun, S. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features," in *Proc. International Conference on Computer Vision*, 2019, pp. 6022–6031.

[109] X. Yao, R. Li, J. Zhang, J. Sun, and C. Zhang, "Explicit Boundary Guided Semi-Push-Pull Contrastive Learning for Supervised Anomaly Detection," in *Proc. Computer Vision and Pattern Recognition*, 2023, pp. 24 490–24 499.

[110] K. Roth, L. Pemula, J. Zepeda, B. Scholkopf, T. Brox, and P. Gehler, "Towards Total Recall in Industrial Anomaly Detection," in *Proc. Computer Vision and Pattern Recognition*, 2022, pp. 14 298–14 308.

[111] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, "MIMII DG: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task," in *Proc. Detection and Classification of Acoustic Scenes and Events 2022 Workshop*, 5 pages, 2022.

[112] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, "ToyADMOS2025: The Evaluation Dataset for the DCASE2025T2 First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring," in *Proc. Detection and Classification of Acoustic Scenes and Events 2025 Workshop*, 2025, pp. 230–234.

[113] T. Nishida, H. Noboru, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Description and Discussion on DCASE 2025 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring," in *Proc. Detection and Classification of Acoustic Scenes and Events 2025 Workshop*, 2025, pp. 55–59.

[114] H. Purohit, T. Nishida, K. Dohi, T. Endo, and Y. Kawaguchi, "MIMII-Agent: Leveraging LLMs with Function Calling for Relative Evaluation of Anomalous Sound Detection," in *Proc. Detection and Classification of Acoustic Scenes and Events 2025 Workshop*, 2025, pp. 160–164.

[115] H. Purohit, R. Tanabe, T. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, "MIMII Dataset: Sound Dataset for Malfunctioning Industrial Ma-

chine Investigation and Inspection," in *Proc. Detection and Classification of Acoustic Scenes and Events 2019 Workshop*, 2019, pp. 209–213.

[116] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, "ToyADMOS: A Dataset of Miniature-machine Operating Sounds for Anomalous Sound Detection," in *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics*, 2019, pp. 308–312.

[117] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[118] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves ImageNet classification," in *Proc. Computer Vision and Pattern Recognition*, 2020, pp. 10 687–10 698.

[119] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *Proc. International Conference on Learning Representations*, 8 pages, 2019.

[120] L. N. Smith and N. Topin, "Super-convergence: very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006, 2019, pp. 369–386.

[121] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

[122] K. Schmid, F. Ritz, S. Illium, and R. Müller, "Analysis of Feature Representations for Anomalous Sound Detection," in *Proc. International Conference on Agents and Artificial Intelligence*, 2021, pp. 97–106.

[123] I. Kuroyanagi, T. Fujimura, K. Takeda, and T. Toda, “Improving anomalous sound detection through pseudo-anomalous set selection and pseudo-label utilization under unlabeled conditions,” *APSIPA Transactions on Signal and Information Processing*, vol. 14, no. 1, 2025, 28 pages.

[124] T. Fujimura, I. Kuroyanagi, and T. Toda, “Discriminative Anomalous Sound Detection Using Pseudo Labels, Target Signal Enhancement, and Ensemble Feature Extractors,” in *Proceedings of the 10th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE 2025)*, 2025, pp. 180–184.

[125] K. Wilkinghoff, “Self-Supervised Learning for Anomalous Sound Detection,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 276–280.

[126] V. Zavrtanik, M. Marolt, M. Kristan, and D. Skočaj, “Anomalous Sound Detection by Feature-Level Anomaly Simulation,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 1466–1470.

[127] Y. Wang, Q. Zhang, Y. Zhang, and J. Hu, “Anomalous Sound Detection Based Feature Fusion and Dual-path Non-linear Independent Components Estimation,” in *Proc. Interspeech*, 2025, pp. 2615–2619.

[128] Y.-Y. Yang, M. Hira, Z. Ni, A. Astafurov, C. Chen, C. Puhrsch, D. Pollack, D. Genzel, D. Greenberg, E. Z. Yang, J. Lian, J. Hwang, J. Chen, P. Goldsborough, S. Narenthiran, S. Watanabe, S. Chintala, and V. Quenneville-Bélair, “Torchaudio: Building Blocks for Audio and Speech Processing,” in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 6982–6986.

[129] J. Hwang, M. Hira, C. Chen, X. Zhang, Z. Ni, G. Sun, P. Ma, R. Huang, V. Pratap, Y. Zhang, A. Kumar, C.-Y. Yu, C. Zhu, C. Liu, J. Kahn, M. Ravanelli, P. Sun, S. Watanabe, Y. Shi, and Y. Tao, "TorchAudio 2.1: Advancing Speech Recognition, Self-Supervised Learning, and Audio Processing Components for Pytorch," in *Proc. Automatic Speech Recognition and Understanding*, 2023, pp. 1–9.

[130] Y. Wang, X. Deng, J. Jiang, and Q. Kong, "ANOMALOUS SOUND DETECTION BASED ON PSEUDO LABELS FROM GUIDED CLUSTERING," DCASE2024 Challenge, Tech. Rep., 2024, 3 pages.

[131] Z. Lv, A. Jiang, B. Han, Y. Liang, Y. Qian, X. Chen, J. Liu, and P. Fan, "AITHU System for First-Shot Unsupervised Anomalous Sound Detection," DCASE2024 Challenge, Tech. Rep., 2024, 4 pages.

[132] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, "ToyADMOS2: Another Dataset of Miniature-Machine Operating Sounds for Anomalous Sound Detection under Domain Shift Conditions," in *Proc. Detection and Classification of Acoustic Scenes and Events 2021 Workshop*, 2021, pp. 1–5.

[133] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, "ToyADMOS2+: New Toyadmos Data and Benchmark Results of the First-Shot Anomalous Sound Event Detection Baseline," in *Proc. Detection and Classification of Acoustic Scenes and Events 2023 Workshop*, 2023, pp. 41–45.

[134] D. Albertini, F. Augusti, K. Esmer, A. Bernardini, and R. Sannino, "IMAD-DS: A Dataset for Industrial Multi-Sensor Anomaly Detection Under Domain Shift Conditions," in *Proc. Detection and Classification of Acoustic Scenes and Events 2024 Workshop*, 2024, pp. 1–5.

[135] D. Niizumi, N. Harada, Y. Ohishi, D. Takeuchi, and M. Yasuda, "ToyADMOS2#: Yet Another Dataset for the DCASE2024 Challenge Task 2 First-Shot Anomalous Sound Detection," in *Proc. Detection and Classification of Acoustic Scenes and Events 2024 Workshop*, 2024, pp. 106–110.

[136] D. K. McClish, "Analyzing a Portion of the ROC Curve," *Medical Decision Making*, vol. 9, no. 3, pp. 190–195, 1989.

[137] J. Ebbers, R. Haeb-Umbach, and R. Serizel, "Threshold Independent Evaluation of Sound Event Detection Scores," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 1021–1025.

[138] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio Set: An ontology and human-labeled dataset for audio events," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 776–780.

[139] I. Kuroyanagi and T. Komatsu, "Self-Supervised Learning Method Using Multiple Sampling Strategies for General-Purpose Audio Representation," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 3263–3267.

[140] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[141] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko, "Quantization and Training of Neural Networks for Efficient

Integer-Arithmetic-Only Inference," in *Proc. Computer Vision and Pattern Recognition*, 2018, pp. 2704–2713.

[142] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization," in *Proc. International Conference on Pattern Recognition*, 2021, pp. 475–489.

[143] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, 2021.

List of Publications

論文誌

[1] I. Kuroyanagi, T. Fujimura, K. Takeda, T. Toda, "Improving Anomalous Sound Detection through Pseudo-anomalous Set Selection and Pseudo-label Utilization under Unlabeled Conditions," *APSIPA Transactions on Signal and Information Processing*, Vol. 14, No. 1, e13, pp. 1–28, 2025.

[2] I. Kuroyanagi, T. Hayashi, K. Takeda, T. Toda, "Serial-OE: Anomalous Sound Detection Based on Serial Method with Outlier Exposure Capable of Using Small Amounts of Anomalous Data for Training," *APSIPA Transactions on Signal and Information Processing*, Vol. 14, No. 1, e1, pp. 1–32, 2025.

国際会議

[3] T. Fujimura, I. Kuroyanagi and T. Toda, "Discriminative Anomalous Sound Detection Using Pseudo Labels, Target Signal Enhancement, and Ensemble Feature Extractors," in *Proc. Detection and Classification of Acoustic Scenes and Events 2025 Workshop*, 2025, pp. 180–184.

[4] T. Fujimura, I. Kuroyanagi, and T. Toda, "Improvements of Discriminative Feature Space Training for Anomalous Sound Detection in Unlabeled Conditions," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2025, pp. 1–5.

[5] I. Kuroyanagi and T. Komatsu, "Self-Supervised Learning Method Using Multiple Sampling Strategies for General-Purpose Audio Representation," in *Proc. International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 3263–3267.

[6] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Improvement of anomalous sound detection method considering the distribution of embedding," in *Proc. International Congress on Acoustics*, 2022, 5 pages.

[7] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Improvement of Serial Approach to Anomalous Sound Detection by Incorporating Two Binary Cross-Entropies for Outlier Exposure," in *Proc. European Signal Processing Conference*, 2022, pp. 294–298.

[8] I. Kuroyanagi, T. Hayashi, Y. Adachi, T. Yoshimura, K. Takeda, and T. Toda, "An Ensemble Approach to Anomalous Sound Detection Based on Conformer-Based Autoencoder and Binary Classifier Incorporated with Metric Learning," in *Proc. Detection and Classification of Acoustic Scenes and Events 2021 Workshop*, 2021, pp. 110–114.

[9] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Anomalous Sound Detection Using a Binary Classification Model and Class Centroids," in *Proc. European Signal Processing Conference*, 2021, pp. 1995–1999.

[10] T. Hayashi, T. Yoshimura, M. Inuzuka, I. Kuroyanagi, O. Segawa, “Spontaneous speech summarization: Transformers all the way through,” in *Proc. European Signal Processing Conference*, 2021, pp. 456–460.

国内会議

[11] 畔柳 伊吹, 林 知樹, 武田 一哉, 戸田 智基, “二種の二値分類タスクに基づく外れ値検出を用いた直列型異常音検知法,” 信学技報, Vol. 122, No. 20, EA2022-8, pp. 35–40, 2022.

[12] 畔柳 伊吹, 林 知樹, 武田 一哉, 戸田 智基, “距離学習を導入した二値分類モデルによる異常音検知,” 音講論, 3-1-15, pp. 277–278, 2021.

[13] 畔柳 伊吹, 林 知樹, 武田 一哉, 戸田 智基, “特徴量空間のクラス重心を考慮した二値分類モデルによる異常音検知,” 信学技報, Vol. 120, No. 397, EA2020-79, pp. 114–121, 2021.

テクニカルレポート

[14] T. Fujimura, I. Kuroyanagi, T. Toda, “The NU systems for DCASE 2025 Challenge Task 2,” Technical report, DCASE Task 2, 5 pages, 2025.

[15] T. Fujimura, I. Kuroyanagi, T. Toda, “The NU systems for DCASE 2024 Challenge Task 2,” Technical report, DCASE Task 2, 5 pages, 2024.

[16] T. Fujimura, I. Kuroyanagi, T. Hayashi, T. Toda, “Anomalous sound detection by end-to-end training of outlier exposure and normalizing flow with domain generalization techniques,” Technical report, DCASE Task 2, 5 pages, 2023.

[17] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, “Two-stage anomalous sound detection systems using domain generalization and specialization techniques,” Technical report, DCASE Task 2, 5 pages, 2022.

[18] I. Kuroyanagi, T. Hayashi, Y. Adachi, T. Yoshimura, K. Takeda, and T. Toda, “Anomalous sound detection with ensemble of autoencoder and binary classification approaches,” Technical report, DCASE Task 2, 5 pages, 2021.

受賞

[19] DCASE 2025 Task2 Judge’s award, 2025

[20] 第 26 回日本音響学会東海支部優秀発表賞, 2022

[21] DCASE 2022 Task2 Judge’s award, 2022

[22] 第 23 回日本音響学会 学生優秀発表賞, 2021