

# 選択的分類による新奇の異常サンプル検知システムの構築

山口 拓海<sup>1,2,a)</sup> 村川 正宏<sup>2,1,b)</sup>

**概要:** 機械学習による異常検知システムでは、異常サンプルの収集は難しいため、正常サンプルのみを用いてモデリングを行うことが一般的であり、入力サンプルが正常か否かを判定し、正常でないサンプルを異常としている。このような、正常サンプルのみでモデリングを行う問題点として、異常の種類を分類できない点が挙げられる。異常検知するだけでなく、異常の種類を判定し、異常の原因推定につなげることは実用上重要である。しかし、原因ごとに異常サンプル収集するにはよりコストが高くなるため、異常分類器を学習することは容易ではない。本研究では、少量サンプルのみで異常分類器を学習するのに加え、学習段階では収集できなかった新奇の異常パターンの検出を、異常検知システムを運用しながら行うための学習手法を提案する。外観検査向けの異常検知データセットである、MVTec AD で実験を行い、提案手法の新奇異常パターン検知の有効性を示した。提案する学習手法により、異常検知に加えて新奇異常サンプル検知を行うことで、異常サンプルの収集を容易にし、継続的な異常分類モデル改善の循環を実現する。

## 1. 序論

近年、生産ラインの自動化に伴い、異常検知システムを用いて、生産ラインの異常を自動的に検知することが一般的となっている。正常サンプルに比べて、異常サンプルは発生確率が低く収集コストが高いため、正常サンプルのみで異常検知システムを構築することが現在の主流となっている。このような正常サンプルのみで異常検知モデルを学習する場合、モデルの判断基準は正常かそうでないかの二択となる。正常サンプルによる異常検知は、様々な手法が提案されており、十分に実用可能なレベルとなっている [1], [2], [3].

本研究では、異常検知に加えて異常の分類を行う。異常分類を行うことで、異常の種類から原因の特定につなげることができる。しかし、上述の正常サンプルのみで学習した異常検知モデルは異常か否か以外判定することができない。そこで、本研究では少量の異常サンプルを用いて、異常分類器を構築する手法を提案する。

また、初期学習時に全ての異常パターンを収集することは難しいと考えられるため、学習済みモデルを稼働しな

がら、新奇異常パターンの検知を行い、異常サンプル収集も効率化する。本研究では、新奇異常パターンを検知するために、Selective Classification [4] の概念を導入する。Selective Classification とは、“Classification with a reject option” を指し、推論の棄却 (Reject) を推論結果の選択肢に追加する。Selective Classification では、分類を間違えるであろう、不確実性が高いサンプルは棄却して推論を行わず、不確実性が低い=信頼度の高いサンプルのみ推論を行うことで、誤った推論を抑制する。Selective Classification の不確実性推定を活用し、異常分類の不確実性を測ることで、新奇異常パターンの検知を行う。

本研究では、異常検知・異常分類・新奇異常パターンの検知を行うために、部分空間法による CNN 特徴量変換、データ拡張による仮想異常データ生成、推論の不確実性推定を異常検知モデルの学習へ導入することを提案する。外観検査向けの異常検知データセットである、MVTec AD [5] で実験を行い、提案手法により異常分類性能の向上、新奇異常パターンの検知性能の向上を達成した。

以下、2 章では本研究の関連研究を示し、3 章では提案手法について、4 章では実験と実験結果について、5 章で結論を述べる。

## 2. 関連研究

### 2.1 部分空間法を用いた異常検知

部分空間法による異常検知 [6] では、大量の正常サンプルを用いて、正常サンプルをよく表現する正常部分空間を求め、正常部分空間へ射影したベクトルと元のベクトルの

<sup>1</sup> 筑波大学大学院  
University of Tsukuba, Tennodai, Tsukuba, Ibaraki 305-8577, Japan

<sup>2</sup> 産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology (AIST), Umezono, Tsukuba, Ibaraki 305-8560, Japan

a) yamaguchi.t@aist.go.jp

b) m.murakawa@aist.go.jp

差分ベクトルのノルムを異常度として、異常検知を行う。部分空間法の利点として、正常部分空間のみを求めればよい。そのため、異常サンプルが無くても、大量の正常サンプルを活用して異常検知性能を高めることが可能である。

$D$ 次元の入力ベクトル  $\mathbf{x} \in \mathbb{R}^D$  のとき、その正常部分空間を張る  $m$  個の  $D$ 次元正規直交ベクトルを  $\mathbf{u}_1, \dots, \mathbf{u}_m$  と表す。正常部分空間への射影行列  $\mathbf{U}$  は、正常部分空間の基底を並べることで得ることができる。

$$\mathbf{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_m\} \quad (1)$$

また、正常部分空間へ射影された  $m$ 次元ベクトルを元の  $D$ 次元空間で見ると、射影行列とその転置の行列積から、 $D \times D$ 変換行列  $\mathbf{P}$  を得る。

$$\mathbf{P} = \mathbf{U}\mathbf{U}^\top \quad (2)$$

元のベクトル  $\mathbf{x}$  を変換行列  $\mathbf{P}$  で、正常部分空間へ射影する。

$$\mathbf{x}' = \mathbf{P}\mathbf{x} \quad (3)$$

元のベクトル  $\mathbf{x}$  と正常部分空間へ射影されたベクトル  $\mathbf{x}'$  の差分  $\tilde{\mathbf{x}}$  を求め、異常度  $a \in \mathbb{R}$  を求める。

$$\tilde{\mathbf{x}} = \mathbf{x}' - \mathbf{x} \quad (4)$$

$$a = \|\tilde{\mathbf{x}}\|_2 \quad (5)$$

正常部分空間の次元数  $m$  の決定方法として、累積寄与率  $\eta \in [0, 1]$  を用いることが一般的である。各基底の固有値  $\lambda_1, \dots, \lambda_D$  より、寄与率の高い順に基底を選択し、一定の累積寄与率になるまで部分空間の基底を増やしていく。

$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^D \lambda_j} \sim \eta \quad (6)$$

## 2.2 データ拡張

データ拡張とは、深層学習モデルの学習において、学習データの増強とも呼ばれる。画像認識の場合、画像の回転や反転を施すことで、見かけ上の学習データ数を増強し、深層学習の汎化性能を向上させる。特に、mixup [7] データ拡張では、2つのサンプルを線形結合することで、学習データに含まれない新しい学習データを生成する。学習サンプル  $\mathbf{x}_i \in \mathbb{R}^D$  と、 $N_c$  クラス分類時の教師ラベル  $\mathbf{y}_i \in [0, 1]^{N_c}$  としたとき、mixup 混合比  $\lambda$  でペアを線形結合し、新たにサンプル  $\hat{\mathbf{x}}$  と教師ラベル  $\hat{\mathbf{y}}$  を付与する。

$$\lambda \in [0, 1] \quad (7)$$

$$\hat{\mathbf{x}} = \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 \quad (8)$$

$$\hat{\mathbf{y}} = \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2 \quad (9)$$

付与された教師ラベルより、分類モデル  $F: \mathbf{x} \in \mathbb{R}^D \rightarrow \mathbf{p} \in [0, 1]^{N_c}$  を Cross Entropy 損失関数で最適化する。

$$\mathbf{p} = \text{SoftMax}(F(\hat{\mathbf{x}})) \quad (10)$$

$$L(\hat{\mathbf{y}}, \mathbf{p}) = - \sum_{i=1}^{N_c} \hat{\mathbf{y}}_i \log p_i \quad (11)$$

## 2.3 Selective Classification

推論時の不確実性の予測値を元に推論結果を信用するかどうかを決定し、誤った推論を抑制する、Selective Classification が提案されている [4]。Selective Classification とは、“Classification with a reject option” を指し、分類器の精度と推論 Coverage のトレードオフを測る問題設定である。推論 Coverage とは、データセット中で推論されるサンプルの割合のことを指し、一般的には、推論 Coverage と分類器の精度にトレードオフの関係が成立する。

分類精度を高めるために、不確実性の低いサンプルのみ推論すると、Coverage は低くなる。逆に、分類精度を犠牲にし、不確実性の高いサンプルも推論すると、Coverage は高くなる。このように、Coverage と精度のトレードオフの関係を利用し、Coverage の変動を許容することで、分類精度の調整を行うことを Selective Classification と言う。

分類モデルを  $F: X \rightarrow Y$ 、選択関数を  $G: X \rightarrow \mathbb{R}$  と定義する。 $X$  は入力空間、 $Y$  は分類ラベル空間。入力サンプル  $\mathbf{x} \in X$  に対し、分類モデル  $F(\mathbf{x})$  は選択関数  $G(\mathbf{x})$  の出力がしきい値  $h$  以上の場合にのみ、推論結果を出力する。

$$(F, G)(\mathbf{x}) = \begin{cases} F(\mathbf{x}) & G(\mathbf{x}) \geq h \\ \text{REJECT} & \text{otherwise} \end{cases} \quad (12)$$

## 3. 提案手法

本研究では、異常検知に加えて、異常分類と新奇異常パターンの検知を行うために、少量の異常サンプルを用いて異常分類器を学習する。少量の異常サンプルのみで分類を行うために、3.1) 部分空間法を用いた特徴量変換を導入し、3.2) mixup により仮想異常データを生成し異常分類を学習する。そして、3.4) Selective Classification による新奇異常パターンの検知を行う。

### 3.1 部分空間法による CNN 特徴量変換

畳み込みニューラルネットワーク (CNN) により画像の特徴抽出を行い、異常検知へ利用する。また、CNN は大規模データセットの ImageNet [8] で事前学習した重みを用いる。異常分類の学習には、少量の異常データしか学習に利用できないため、一般的な最適化では分類の学習は難しい。そこで本研究では、大量に利用できる正常サンプルから、正常部分空間を構築し、異常検知で用いられていた、射影ベクトルと元のベクトルの差分ベクトル  $\tilde{\mathbf{x}}$  (式 (4)) を異常分類にも活用する。

本研究では、差分ベクトル  $\tilde{\mathbf{x}}$  の長さが異常度を示し、方向が異常の種類を示すと仮定する。異常度を差分ベクトルの L2 ノルムから推定し、異常分類は差分ベクトル

を L2 ノルムで正規化し、超球表面上に写すことで、ベクトルの方向から異常分類を学習する。Channel  $C$ , Height  $H$ , Width  $W$  の画像を  $D$  次元特徴量へ変換するモデルを  $F_E: \mathbf{x} \in \mathbb{R}^{CHW} \rightarrow \mathbf{z} \in \mathbb{R}^D$  と定義する。

$$\mathbf{z} = F_E(\mathbf{x}) \quad (13)$$

正常部分空間への変換行列である  $\mathbf{P}$  (式 (2)) より、特徴量を正常部分空間へ写像し、元のベクトルとの差分  $\tilde{\mathbf{z}}$  を求める。

$$\mathbf{z}' = \mathbf{P}\mathbf{z} \quad (14)$$

$$\tilde{\mathbf{z}} = \mathbf{z}' - \mathbf{z} \quad (15)$$

$D$  次元特徴量を  $N_c$  クラス分類するモデルを  $F_C: \mathbf{z} \in \mathbb{R}^D \rightarrow \mathbf{p} \in [0, 1]^{N_c}$  とし、異常度  $a \in \mathbb{R}$  および 異常分類結果  $\mathbf{p} \in [0, 1]^{N_c}$  を求める。

$$a = \|\tilde{\mathbf{z}}\|_2 \quad (16)$$

$$\mathbf{p} = \text{SoftMax}(F_C(\tilde{\mathbf{z}}/a)) \quad (17)$$

このように、正常部分空間を用いて少量の異常サンプルから分類器を学習し、異常検知は従来と同様の方法で行う。

### 3.2 Mixup による仮想異常データ生成

異常分類の学習に使用できる異常データが非常に少量であるため、分類器の汎化性能向上のためにデータ拡張の mixup [7] を用いる。画像ピクセル単位での異常データ合成は難しいため、本研究では mixup は入力空間（画像ピクセル）ではなく、中間特徴 (式 (13)) 空間で行う。また、以下 2 種類の mixup により、異常サンプルを水増しする。

#### 3.2.1 正常 + 異常サンプル

正常サンプル  $\mathbf{z}_g$  と異常サンプルを  $\mathbf{z}_a$  を線形結合し、仮想の異常サンプル  $\hat{\mathbf{z}}$  を得る。教師ラベル  $\hat{\mathbf{y}}$  は、合成した異常サンプルの異常クラス  $\mathbf{y}_a \in [0, 1]^{N_c}$  を付与する。

$$\hat{\mathbf{z}} = \lambda\mathbf{z}_g + (1 - \lambda)\mathbf{z}_a \quad (18)$$

$$\hat{\mathbf{y}} = \mathbf{y}_a \quad (19)$$

#### 3.2.2 異常 + 異常サンプル

異なる異常サンプル同士 ( $\mathbf{z}_i, \mathbf{y}_i$ ), ( $\mathbf{z}_j, \mathbf{y}_j$ ) ( $i \neq j$ ) を線形結合し、仮想異常サンプル  $\hat{\mathbf{z}}$  を得る。教師ラベルは、ソフトラベル  $\hat{\mathbf{y}} \in [0, 1]^{N_c}$  を得る。

$$\hat{\mathbf{z}} = \lambda\mathbf{z}_i + (1 - \lambda)\mathbf{z}_j \quad (20)$$

$$\hat{\mathbf{y}} = \lambda\mathbf{y}_i + (1 - \lambda)\mathbf{y}_j \quad (21)$$

### 3.3 中間特徴を用いた mixup

本研究では、学習済み畳み込みニューラルネットワーク (CNN) の高い表現能力を活用し、中間特徴量による mixup データ拡張を行う。中間特徴量による mixup により、入力空間よりも高品質なデータ拡張が可能となる [9]。図 1

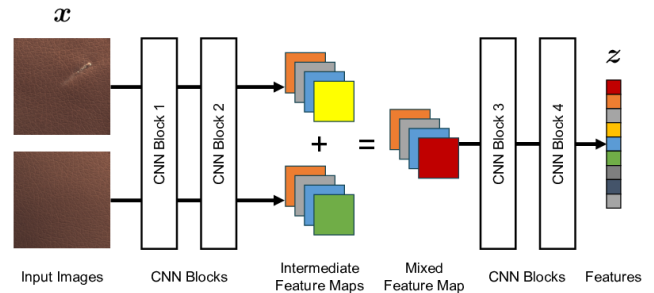


図 1: 中間特徴を用いた mixup

に示すように、CNN の中間特徴を抽出し、一定の比率で mixup を行う。mixup で生成された特徴量を、後段の畳み込み層で更に変換し、最終的な特徴量ベクトルを得る。

### 3.4 Selective Classification による新奇異常パターンの検知

異常分類が難しいサンプルが新奇異常パターンであると考えられる。本研究では、異常分類器の出力のエントロピーより、新奇異常度  $a_n \in \mathbb{R}$  を求める。

$$a_n = - \sum_{j=1}^{N_c} p_j \log p_j \quad (22)$$

## 4. 実験・実験結果

提案手法の有効性を確かめるため、一般的な異常検知タスクに加えて、異常分類と新奇異常パターンの検知性能を評価する。

### 4.1 データセット

外観検査のデータセットである、MVTec AD [5] より、本実験では leather, carpet で検証する。表 1, 表 2, 図 2, 図 3 にデータセット詳細を示す。

異常分類の学習のために、Test データの各異常クラスから約 4 枚ずつ学習へ使用する。また、ハイパラメータ探索のため、Test データの 40% を Validation へ利用する。これらの学習やハイパラメータ探索で使用したサンプルは Test から取り除く。

### 4.2 評価方法

leather, carpet とともに、異常クラスが 5 クラスずつ割り振られている。新奇異常パターンの検知性能を測るため、異常クラスから 1 つ新奇異常クラスを選択し、残りの 4 クラス分類を学習する。評価時に新奇異常パターンのサンプルをモデルで推論し、新奇異常パターン検知性能を ROCAUC, PRAUC で評価する。

各異常クラスについて全て実験を行い、その調和平均で評価値を集約する。

#### 4.2.1 異常検知

新奇異常パターンも含めて、異常検知性能の ROCAUC・

PRAUC を評価する．よって，学習していない異常パターンも異常検知性能の評価で用いる．

#### 4.2.2 異常分類

異常サンプルの異常分類精度を評価する．正常サンプルは評価から除外する．

#### 4.2.3 新奇異常パターン検知

異常サンプルの中で，新奇パターンか学習済みの異常かの2値分類を ROCAUC・PRAUC で評価する．正常サンプルは評価から除外する．

### 4.3 比較手法

比較のベースラインとして，正常クラス・異常クラス両方を分類するモデルを構築する．つまり，分類モデルは入力サンプルを Good, 異常クラス1, 異常クラス2, ... のいずれかへ分類する．また，部分空間法による特徴量変換や mixup によるデータ拡張は行わない．提案手法と同様に大量の正常サンプルと約4枚の異常サンプルで分類器を学習する．

異常クラス数を  $N_c$  としたとき，特徴抽出モデルは  $F_E: \mathbf{x} \in \mathbb{R}^{CHW} \rightarrow \mathbf{z} \in \mathbb{R}^D$ ，分類モデルは  $F_C: \mathbf{z} \in \mathbb{R}^D \rightarrow \mathbf{p} \in [0, 1]^{N_c+1}$  と定義できる．異常度  $a \in \mathbb{R}$  は，推論結果の正常クラスへの所属確率を符号反転させることで求める．

$$\mathbf{z} = F_E(\mathbf{x}) \quad (23)$$

$$\mathbf{p} = \text{SoftMax}(F_C(\mathbf{z})) \quad (24)$$

$$a = -\mathbf{p}_{\text{good}} \quad (25)$$

また，新奇異常度  $a_n$  は，正常クラスも含めてエントロピーを計算する．

$$a_n = -\sum_{j=1}^{N_c+1} \mathbf{p}_j \log \mathbf{p}_j \quad (26)$$

#### 4.4 使用モデル・パラメータ

- 特徴抽出モデル: 事前学習済み EfficientNet-B4 [10]
  - 分類モデル: 1層 Linear
  - Optimizer: MADGRAD [11]
  - LearningRate: 1e-3
  - WeightDecay: 0
  - Loss: CrossEntropy Loss + Class Imbalance 補正
- 特徴抽出モデルの重みは固定し，学習しない．他，以下のハイパラメータについて，Validation データを用いて最適値を探索する．
- 正常部分空間の累積寄与率: {0.9, 0.99, 0.999, 0.9999} から1つ選択する．
  - Mixup を行う中間層の深さ: 特徴量抽出モデルから深い順に4層抽出し，その中で1つ選択する．

表 1: MVTec AD leather サンプル数

Classes		Train	Test
Good		245	32
Anomaly	fold	0	17
	color	0	19
	cut	0	19
	glue	0	19
	poke	0	18

表 2: MVTec AD carpet サンプル数

Classes		Train	Test
Good		280	28
Anomaly	color	0	19
	thread	0	19
	cut	0	17
	hole	0	17
	metal_contamination	0	17

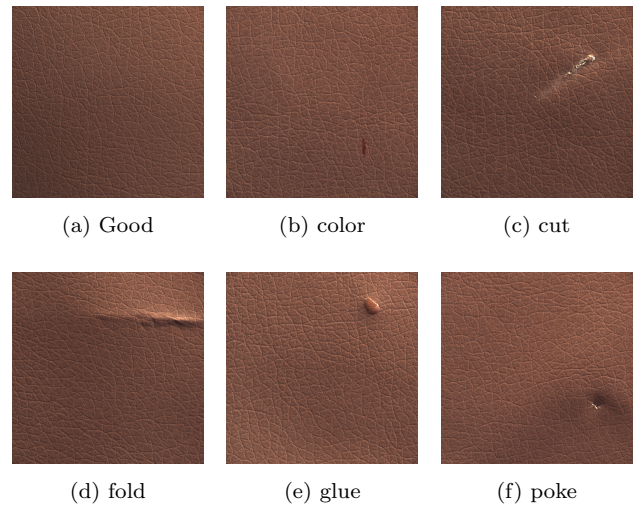


図 2: MVTec AD leather

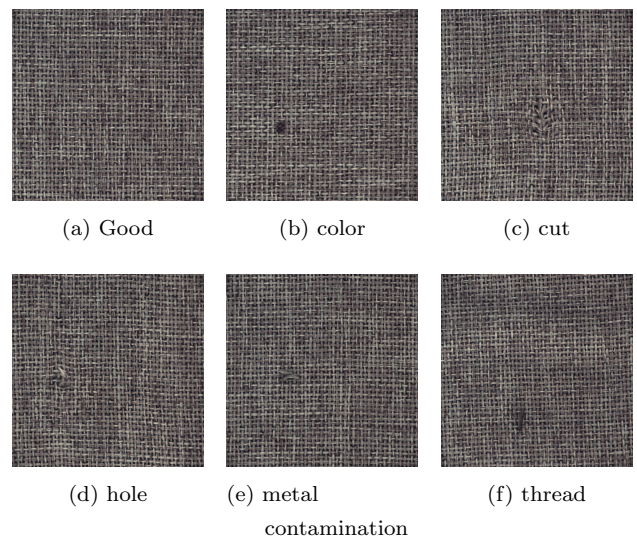


図 3: MVTec AD carpet

表 3: 異常検知・新奇異常パターン検知性能 (leather)

Method	Subspace	Mixup	Anomaly Detection	Anomaly Detection			Novel Anomaly Detection	
				ROCAUC	PRAUC	Classification Acc.	ROCAUC	PRAUC
Baseline			Classifier	0.997	0.999	0.456	0.472	0.319
Ours	✓		Subspace	<b>1.000</b>	<b>1.000</b>	0.870	0.667	0.370
Ours	✓	✓	Subspace	<b>1.000</b>	<b>1.000</b>	<b>0.931</b>	<b>0.768</b>	<b>0.465</b>

表 4: 異常検知・新奇異常パターン検知性能 (carpet)

Method	Subspace	Mixup	Anomaly Detection	Anomaly Detection			Novel Anomaly Detection	
				ROCAUC	PRAUC	Classification Acc.	ROCAUC	PRAUC
Baseline			Classifier	0.999	<b>1.000</b>	0.598	0.729	0.418
Ours	✓		Subspace	<b>1.000</b>	<b>1.000</b>	<b>0.792</b>	0.729	0.438
Ours	✓	✓	Subspace	<b>1.000</b>	<b>1.000</b>	0.780	<b>0.749</b>	<b>0.534</b>

#### 4.5 実験結果

表 3, 表 4 に実験結果を示す. leather の場合, 部分空間法と mixup を組み合わせることで, 異常検知・異常分類・新奇異常パターン検知, 全ての性能が向上した. carpet の場合, 異常分類を除いて, 部分空間法と mixup の組み合わせで, 性能が向上していることが確認できる.

部分空間法・mixup 共に, 異常分類性能の向上と新奇異常パターン検知性能の向上に寄与している. 特に leather の場合, ベースラインの分類精度 45.6% に対し, 部分空間法による特徴量変換を導入することで 87.0%, 更に mixup を導入することで 93.1% まで分類性能を向上させることに成功した. このことから, 提案手法により, 学習へ利用できる異常サンプルが少量であっても, 異常分類・新奇異常パターン検知ができていくことがわかる.

続いて, 実運用を考慮した評価として, Coverage ごとの異常検知性能を表 5, 表 6 に示す. この評価では, 異常分類の推論不確実性が高い順 (新奇異常パターンである可能性が高い順) に, 推論器の判定結果を示す. Coverage が低いほど, 推論の不確実性は高くなるため, 異常分類が間違っている場合や, 新奇異常パターンである可能性が高い.

いずれのクラスの結果でも, 低 Coverage (推論の不確実性が高い) のサンプルは, 半分程度が新奇異常パターンとなり, 本研究で提案している不確実性をベースとした新奇異常パターン検知に成功している. また, 低 Coverage のサンプルは, 異常分類性能も低くなっていることから, 学習済みの異常パターンについても, 曖昧なサンプルの検知が可能であることが確認できる.

提案手法により異常検知・異常分類・新奇異常パターン検知が可能であることを確認した. また, 異常分類が曖昧な結果となったサンプルの検出も可能であることを確認した.

#### 4.6 考察

##### 4.6.1 異常検知の不確実性推定

異常検知システムの実稼働では, サンプル収集やアンノテーションのコストが問題となる. 本手法では異常検知システムに不確実性推定機構を導入することにより, 確実な異常サンプルは完全自動で処理し, 不確実な異常のみ, 人間による再点検やデータ蓄積へ役立てることが可能となる. 不確実性の高いサンプルのみ抽出し, 人間による再点検の工数を必要最低限にすることでデータ収集コストを抑え, 継続的な異常検知・異常分類モデルの改善を実現できると考えられる.

##### 4.6.2 新奇異常パターンによる性能のばらつき

carpet の結果 (表 4, 表 6) では, leather の結果と比較して, mixup の効果が十分に確認できなかった. また, 新奇異常パターンを poke としたとき, 推論の不確実性に基づく新奇異常パターン検知に失敗している (表 6 poke).

この現象は, ハイパラメータである正常部分空間の累積寄与率の設定や, mixup を行う中間層の深さが原因であると考えられる. 目的としている, 新奇異常パターン検知モデルは, 様々なパターンの新奇異常を検知しなければならない. しかし, 新奇異常の種類によって, 最適な正常部分空間の累積寄与率や, mixup の深さが異なる場合が考えられる. 本実験の場合, poke を検出するために最適なハイパラメータと, 他異常クラスを検出するために最適なハイパラメータが異なることが原因として考えられる.

この課題を解決するためには, 特徴抽出モデルの追加学習や mixup への工夫が必要であると考えられ, 今後の課題としたい.

## 5. 結論

異常検知システムの実運用において, 異常検知だけでなく, 異常原因の特定のために異常の種類判別も重要である. 本研究では, 異常分類が可能な異常検知システムの学習について検討した. 異常分類を学習する際の問題点と

表 5: 部分空間法+mixup / Coverage ごとの推論結果 (leather)

Coverage  $a$ - $b$ : 推論の 不確実性を高い順にサンプルを並べ替え, 上位  $a$  割から上位  $b$  割までのサンプルを指す.

Novel rate は, Coverage 中での新奇異常パターンの割合を示す. Classification Acc. は, Coverage 中での異常分類性能を示す.

Novel Anomaly Coverage	color		cut		hole		metal_contamination		thread	
	Novel rate	Classification Acc.	Novel rate	Classification Acc.	Novel rate	Classification Acc.	Novel rate	Classification Acc.	Novel rate	Classification Acc.
0.0-0.2	<b>0.444</b>	<b>0.400</b>	<b>0.556</b>	<b>0.250</b>	<b>0.444</b>	<b>1.000</b>	<b>0.444</b>	<b>0.800</b>	<b>0.444</b>	<b>0.400</b>
0.2-0.4	<b>0.333</b>	<b>0.333</b>	<b>0.556</b>	<b>0.750</b>	<b>0.556</b>	<b>0.500</b>	<b>0.111</b>	<b>1.000</b>	<b>0.111</b>	<b>0.750</b>
0.4-0.6	0.111	0.625	0.000	0.556	0.111	0.750	0.444	0.400	0.222	1.000
0.6-0.8	0.333	0.833	0.000	0.778	0.000	0.778	0.111	1.000	0.222	1.000
0.8-1.0	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.222	1.000

表 6: 部分空間法+mixup / Coverage ごとの推論結果 (carpet)

Coverage  $a$ - $b$ : 推論の 不確実性を高い順にサンプルを並べ替え, 上位  $a$  割から上位  $b$  割までのサンプルを指す.

Novel rate は, Coverage 中での新奇異常パターンの割合を示す. Classification Acc. は, Coverage 中での異常分類性能を示す.

Novel Anomaly Coverage	color		cut		fold		glue		poke	
	Novel rate	Classification Acc.	Novel rate	Classification Acc.	Novel rate	Classification Acc.	Novel rate	Classification Acc.	Novel rate	Classification Acc.
0.0-0.2	<b>0.444</b>	<b>0.600</b>	<b>0.778</b>	<b>1.000</b>	<b>0.444</b>	<b>0.800</b>	<b>0.778</b>	<b>1.000</b>	0.000	1.000
0.2-0.4	<b>0.333</b>	<b>0.667</b>	<b>0.444</b>	<b>0.600</b>	<b>0.444</b>	<b>1.000</b>	<b>0.444</b>	<b>1.000</b>	0.667	1.000
0.4-0.6	0.222	1.000	0.000	1.000	0.200	0.875	0.000	0.778	0.200	1.000
0.6-0.8	0.111	1.000	0.000	1.000	0.000	0.778	0.000	1.000	0.111	1.000
0.8-1.0	0.100	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.200	1.000

して, 異常サンプルの収集は難しく, モデルを学習するのに十分なサンプル数を用意できない. そこで本研究では, 少量の異常サンプルのみで異常分類器を学習するために, 部分空間法による正常部分空間を活用した特徴量変換や, mixup による異常データのデータ拡張の活用を提案した. また, Human-in-the-loop による継続的なモデル改善のために, 未知の異常パターンである, 新奇異常パターンを検知し, 異常サンプルのデータ収集の効率化も提案した.

実世界の外観検査データセットである, MVTec AD で提案手法を検証し, 学習に少量の異常サンプルしか使えない場合でも, 異常分類性能を向上させることに成功した. また, 異常分類器に推論の不確実性推定を導入することで, 新奇異常パターンを検知することに成功した. 更に, 新奇異常パターンだけでなく, 異常種類の判別の難しい曖昧な異常サンプルの検知にも成功した.

今後, Human-in-the-loop を前提とした異常検知システムの研究開発をさらに進め, データ収集の効率化と継続的なモデル改善の両立を実現したい.

**謝辞** この成果の一部は, 国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務 (JPNP20006) の結果得られたものです.

## 参考文献

[1] Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R. and Wu, L.: FastFlow: Unsupervised Anomaly Detection and Localization via 2D Normalizing Flows, *arXiv preprint arXiv:2111.07677* (2021).

[2] Defard, T., Setkov, A., Loesch, A. and Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization, *International Conference on Pattern Recognition*, Springer, pp. 475–489 (2021).

[3] Rippel, O., Mertens, P. and Merhof, D.: Modeling the distribution of normal data in pre-trained deep features for anomaly detection, *2020 25th International Confer-*

*ence on Pattern Recognition (ICPR)*, IEEE, pp. 6726–6733 (2021).

[4] El-Yaniv, R. et al.: On the Foundations of Noise-free Selective Classification., *Journal of Machine Learning Research*, Vol. 11, No. 5 (2010).

[5] Bergmann, P., Fauser, M., Sattlegger, D. and Steger, C.: MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9592–9600 (2019).

[6] Nanri, T. and Otsu, N.: Unsupervised Abnormality Detection in Video Surveillance., *MVA*, pp. 574–577 (2005).

[7] Zhang, H., Cisse, M., Dauphin, Y. N. and Lopez-Paz, D.: mixup: Beyond empirical risk minimization, *arXiv preprint arXiv:1710.09412* (2017).

[8] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: Imagenet: A large-scale hierarchical image database, *2009 IEEE conference on computer vision and pattern recognition*, Ieee, pp. 248–255 (2009).

[9] Verma, V., Lamb, A., Beckham, C., Najafi, A., Mitliagkas, I., Lopez-Paz, D. and Bengio, Y.: Manifold Mixup: Better Representations by Interpolating Hidden States, *Proceedings of the 36th International Conference on Machine Learning (Chaudhuri, K. and Salakhutdinov, R., eds.)*, Proceedings of Machine Learning Research, Vol. 97, PMLR, pp. 6438–6447 (online), available from <https://proceedings.mlr.press/v97/verma19a.html> (2019).

[10] Tan, M. and Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks, *International Conference on Machine Learning*, PMLR, pp. 6105–6114 (2019).

[11] Defazio, A. and Jelassi, S.: Adaptivity without compromise: a momentumized, adaptive, dual averaged gradient method for stochastic optimization, *arXiv preprint arXiv:2101.11075* (2021).