The 39th Annual Conference of the Japanese Society for Artificial Intelligence, 2025

撮像環境が統一できない異常検知タスクに関する一考察

A study of anomaly detection tasks in which the imaging environment cannot be unified

藤井 純一郎[*1] 都築 幸乃[*1] 七海 友康[*1] 大久保 順一[*1]
Junichiro Fujii Yukino Tsuzuki Tomoyasu Nanaumi Junichi Okubo

[*1] 八千代エンジニヤリング株式会社
Yachiyo Engineering, Co., Ltd.

In the field of image anomaly detection tasks, benchmark data such as MVTec AD is widely used. These are datasets from the manufacturing industry, and the imaging environment is uniform. On the other hand, in anomaly detection for outdoor infrastructures, the imaging environment cannot be unified. In this study, we propose a dataset of outdoor concrete blocks (Gogan dataset) which includes a variety of imaging environments. Then, we applied existing anomaly detection methods to the MVTec AD and Gogan datasets under the same conditions and compared their performance. Based on the results, we discuss the issues involved in anomaly detection under conditions where the imaging environment cannot be unified.

1. はじめに

画像の異常検知タスクにおいては，MVTec AD[Bergmann 2019]や VisA[Zou 2022]といったベンチマークデータが広く用いられている．これらは製造業でのデータセットであり，撮像環境が統一されている．一方，社会インフラや屋外構造物での異常検知においては撮像環境が統一できないことが通常である．

本研究では，多様な撮像環境を含む屋外のコンクリートブロックのデータセット（Gogan データセット）を提案する．その上で MVTec AD と Gogan データセットに対して，既存の異常検知手法を同じ条件で適用し，データセットによる異常検知性能の比較実験を行う．実験結果に基づき，撮像環境が統一できない条件での異常検知に向けた課題について考察する．

2. 関連研究・データセット

2.1 既存手法

異常検知タスクにおいて，正常データは大量に取得できるのに対して，異常データは稀で特徴も多様であることから，正常データのみを用いて異常検知を行う教師なし異常検知の問題が研究されてきた．特に深層学習の登場以降は，深層学習モデルで獲得した画像特徴量を用いて教師なし異常検知を行う手法が数多く提案されている．例えば PatchCore[Roth 2022]は，入力した特徴量に対してメモリバンクに保存された正常画像特徴量の中から最近傍特徴量を選択し，その近傍距離を異常スコアとして判定を行う．異常スコアはパッチごとに得られるため，パッチレベルで異常を判定することができる．EfficientAD[Batzner 2024]は蒸留した特徴量抽出器を Teacher とする Student-Teacher モデルにより構造的な異常を抽出し，さらにそれを AutoEncoder と組み合わせることで論理的な異常にも対応した．MSFlow[Zhou 2024]は特徴量抽出器の複数の中間層の正常画像特徴量を正規化フローによりガウス分布に従う潜在変数に変換し，潜在分布上での確率密度が低いデータを異常とみなす．

一方，より厳しい問題設定として正常データの分布がドリフトする条件下で異常検知を行うことを想定し，同一ドメインの正常データを利用しない Zero-shot 異常検知の研究が近年進んでいる．例えば WinCLIP[Jeong 2023]は，CLIP の汎用的な image-text のアライメントを利用して異常検知を実現した．異常検知のためのテキストプロンプトは，対象物の状態などを表す単語を事前に人手で設定し，それらを組み合わせて生成している．AnomalyCLIP[Zhou 2024]も同様に CLIP を用いた異常検知手法であるが，テキストエンコーダに学習可能パラメータを追加し，ターゲットドメインと別ドメインの正常/異常データを用いて最適化することで，人手によるプロンプト調整を回避している．MuSc[Li 2024]は Vit を特徴量抽出器として用い，パッチトークン単位でテストセット内の類似パッチの分布から異常スコアを算出する．

2.2 既存データセット

画像異常検知のデータセットとしては，MVTec AD がベンチマークとして多く用いられ，2.1 に示したいずれの既存手法においても MVTec AD での精度検証が行われている．MVTec AD は製造業における外観検査に焦点を当てた異常検出手法のベンチマーク用データセットで，15 種類のオブジェクトとテクスチャカテゴリに分類された 5,354 枚の画像が含まれている．各カテゴリは，正常なトレーニング画像セットと，様々な種類の異常画像と正常画像が混在するテストセットから構成されている．画像レベルの正常/異常のラベルに加え，ピクセルレベルの異常領域のラベルも含まれる．MVTec AD データセットは構造的な異常データのみを含むのに対し，後に論理的な異常を含む MVTec LOCO AD データセットも追加された．

VisA データセットも画像異常検知で広く用いられている．VisA も製造業における外観検査のデータセットで，12種類のオブジェクトカテゴリに分けられた 10,821 枚で構成される．

これらのデータセットは製造業のデータであり，オブジェクトやテクスチャといったカテゴリの多様性を重視しているのに対し，カテゴリ内の画像は撮像環境が統一されている．データドリフトや製造業以外の用途に対応するためには，撮像環境の多様性を持ったベンチマークデータで検証することが望ましい．撮像環境が統一されていないベンチマークデータとしては，AeBAD[Zhang 2023]データセットがある．AeBAD は航空エンジンブレードの画像データセットで，同一ドメイン内の多様性を持たせるために画角などが統一されていない．さらにドメインシフトを想定し，テストセットでは背景色・照明・画角の分布が異なるデータが用意されている．ただし，AeBAD も背景色・照明・画角を制御できる室内実験で作成されており，屋外環境で生じるような多様性は持っていない．

連絡先：藤井純一郎，八千代ｴﾝｼﾞﾆﾔﾘﾝｸﾞ（株），〒111-8648
東京都台東区浅草橋 5-20-8，jn-fujii@yachiyo-eng.co.jp

3. Gogan データセットの作成

3.1 概要

撮像環境を統一できない屋外の実環境での画像異常検知のためのベンチマークデータとして，河川護岸のコンクリートブロックのデータセット（Gogan データセット）を作成した．対岸から護岸に正対するように撮影した画像から，コンクリートブロックを切り出し，目視によるラベリングを行った．

画像データ仕様やデータセット構成は，後述する実験で MVTec AD と横並びで比較するため，極力 MVTec AD と共通とした．

3.2 画像データ仕様

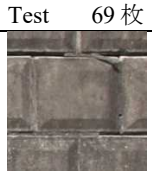
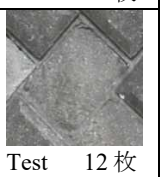
元画像は 5616×3744pixel，72dpi であるが，対岸との距離やアングルにより，個別のブロックを切り出した画像サイズはばらつきが出る．そのため，すべてのブロック画像は 256×256pixel にリサイズした．

画像レベルでの正常/異常に加えて，ピクセルレベルで精度評価を行うため，異常領域をマスクした 2 値画像も作成した．

3.3 データセット構成

ブロック画像はコンクリートブロックの種類により，布積みと谷積みの2カテゴリに分けられている．両カテゴリ内に正常画像と異常画像が含まれ，異常画像は変状種類によりクラス分けされている．データセットの構成とサンプル画像を表 1 に示す．

表 1 Gogan データセット

クラス		布積み	谷積み
正常		 Train 2031 枚 Test 69 枚	 Train 2608 枚 Test 129 枚
異常	ひび割れ/目地開き	 Test 12 枚	 Test 12 枚
	剥離/欠損/外れ	 Test 12 枚	 Test 12 枚
	エフロレッセンス	 Test 12 枚	 Test 12 枚
	溶出物	 Test 12 枚	（データなし）
合計枚数	Train	2031 枚	2608 枚
	Test	117 枚	165 枚

4. データセットによる異常検知精度の比較実験

撮像環境を統一できる屋内環境と，そうでない屋外環境で，異常検知精度にどのような差が出るのかを明らかにするため，MVTec AD データセットと Gogan データセットに対して，既存の異常検知手法を同じ条件で適用し，データセットによる精度を比較する実験を行った．

4.1 対象手法

撮像環境を統一できない異常検知タスクを考えた場合，多様な撮像環境の正常データを用意する教師なし異常検知と，特定の撮像環境のみ用意する few-shot/zero-shot 異常検知の，2 通りの問題設定が考えられる．本研究では教師なし異常検知手法，zero-shot 異常検知手法に分けて，2.1 に挙げた近年 MVTec AD データセットで高い精度を達成したと報告されている下記の手法を対象とした．

(1)教師なし異常検知
- PatchCore
- EfficientAD
- MSFlow

(2)zero-shot 異常検知
- WinCLIP
- AnomalyCLIP
- MuSc

4.2 評価指標

MVTec AD で用いられる下記の評価指標を，Gogan データセットでも同様に算出した．なお，AUPRO についてはピクセルレベルのみでの評価となる．
- AUROC: Area Under ROC curve
- F1: F1 Score
- AP: Average Precision
- AUPRO: Area Under PRO curve

4.3 評価結果

各手法の評価結果を表 2 に示す．また，撮像環境の影響を見るため，各手法の MVTec AD と Gogan のデータセットでの評価結果比較のグラフを図 1 に示す．

まず教師なし異常検知手法について見ると，MVTec AD では画像レベルでは PatchCore が，ピクセルレベルでは MSFlow が最も高い精度を記録したが，Gogan では大きく順位が入れ替わった．MSFlow は他手法に比べて MVTec AD に対する Gogan での精度低下の度合いが大きかった．また，全手法共通で画像レベルよりもピクセルレベルの評価指標の精度低下が大きかった．

次に，Zero-shot 異常検知手法について見ると，MVTec AD では画像レベルでは MuSc が全評価指標で最も高い精度を記録したが，Gogan では 2 指標を除き AnomalyCLIP が最も高い精度であった．MuSc は MVTec AD と Gogan で精度の差が大きく，特にピクセルレベルの F1 と AP においては大幅にスコアが低下している．MuSc を除く 2 手法においては，逆にピクセルレベルの精度低下の度合いが教師なし異常検知の手法と比べても小さく抑えられている．

タスクの難易度の違いから，総じて教師なし異常検知のスコアが zero-shot 異常検知のスコアを上回っているが，Gogan のピクセルレベルの指標では CLIP を用いた手法（WinCLIP，AnomalyCLIP）において一部逆転が起きている．

表 2 精度評価結果

異常検知手法		MVTecAD							Gogan						
		image-level			pixel-level				image-level			pixel-level			
		AUROC	F1	AP	AUROC	F1	AP	AUPRO	AUROC	F1	AP	AUROC	F1	AP	AUPRO
教師なし	PatchCore	**99.02**	**98.53**	**99.67**	98.42	61.73	61.15	93.56	**98.92**	88.04	93.80	**91.76**	**47.96**	**45.15**	70.54
	Efficient AD	98.48	97.80	99.44	96.47	65.06	64.30	88.49	97.06	**89.86**	**95.62**	84.48	46.55	41.68	64.70
	MSFlow	96.89	97.02	98.47	**98.78**	**65.16**	**65.18**	**95.57**	96.66	86.80	93.41	87.63	45.37	40.75	**75.77**
	平均	98.13	97.78	99.19	97.89	63.98	63.54	92.54	97.55	88.23	94.27	87.96	46.63	42.53	70.34
zero-shot	WinCLIP	90.38	92.68	95.64	82.29	24.78	18.17	61.92	87.53	70.93	76.25	78.71	24.81	15.62	49.89
	AnomalyCLIP	91.55	92.75	96.36	91.07	39.08	34.54	81.38	**88.91**	**75.96**	**80.93**	87.26	**41.21**	**36.10**	69.15
	MuSc	**95.67**	**96.41**	**98.10**	**97.29**	**60.71**	**61.29**	**92.14**	84.15	65.00	70.10	**90.25**	1.90	0.45	**82.69**
	平均	92.53	93.95	96.70	90.22	41.53	38.00	78.48	86.86	70.63	75.76	85.40	22.64	17.39	67.24

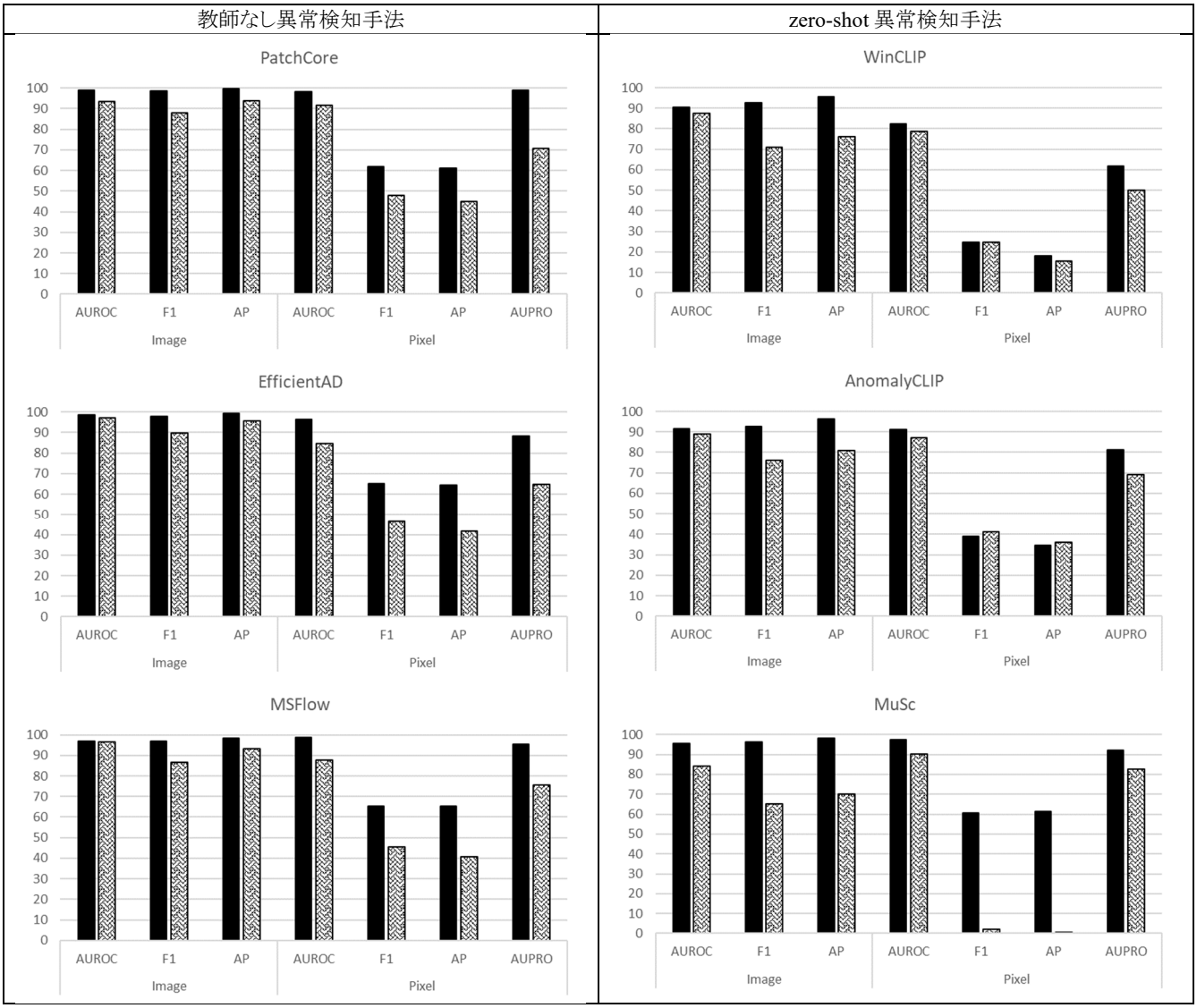


図 1 データセットによる精度評価結果の比較(■MVTec ▨Gogan)

4.4 考察

(1)全体

撮像環境が統一されているMVTec ADとされていないGoganで精度が大きく異なり，かつ一律に精度が低下するのではなく評価指標によって傾向が異なることから，屋外環境などで画像異常検知を適用するためには MVTec AD のような撮像環境が統一されたベンチマークデータの評価では不十分なことが示唆される．

本研究で作成した Gogan データセットにより，撮像環境の多様性を考慮した精度評価が可能となった．ただし，実環境ではより多様な撮像環境が発生し得るため，それらのバリエーションを確保したベンチマークデータの拡充は今後の課題である．

(2)教師なし異常検知

屋外で撮影した Gogan データセットでは撮像環境はコントロールできないため，Train セットと Test セットの撮像環境のばらつきは完全には一致しない．教師なし異常検知手法は，Train セットに含まれる正常画像を学習するため，Train と Test の撮像環境

が統一されている MVTec AD に対して Gogan での精度が低下するのは原理的に必然と考えた.

教師なし異常検知手法に共通して, Gogan では画像レベルに比べてピクセルレベルの評価指標の低下度合いが大きい事象について調査するため, Test セットの Anomaly Map を可視化した例を図 2 に示す. Gogan の正常画像には図 2 に示すようなブロックの一部分が湿っていたり影になっている画像が含まれ, ピクセルレベルで大幅な過検出が生じ得る. これにより, ピクセルレベルの精度評価指標が低下することを確認した.

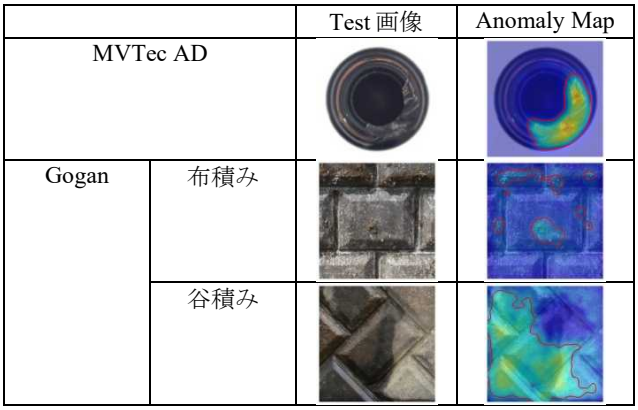


図 2 PatchCore の Anomaly Map 例

(3)zero-shot 異常検知

本研究で比較した 3 手法のうち, Gogan での精度評価結果が MuSc のみ傾向が大きく異なる理由について考察を行った. MuSc は画像内の正常部分のパッチは, Test セット内に類似したパッチが存在する仮定に基づいて異常を判定する. しかし Gogan は撮像環境にばらつきがあるため, 正常部分のパッチにおいても類似したパッチが存在しない場合があるため, 上述の仮定が成り立たない. これによりパッチ単位の異常スコアが過大に算定され, 結果としてピクセルレベルの精度評価が低下したと考えた. Test セットの Anomaly Map を可視化した結果を図 3 に示す. Gogan には図 3 に示すように直射日光が当たって全体に白っぽいブロック画像や, 逆に日向と日影が混在したために暗いブロック画像などがあり, それらは正常であるにも関わらず異常スコアが全体的に高くなっていることが確認できる.

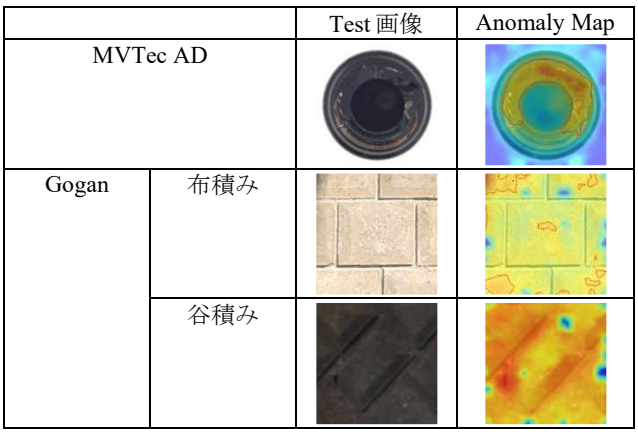


図 3 MuSc の Anomaly Map 例

これは図 2 に示した教師なし異常検知手法でピクセルレベルの精度低下と同じ原因であるが, MuSc の機構上より顕著にその影響が表れている. 一方, WinCLIP と AnomalyCLIP は CLIP の画像特徴量を使っており, CLIP が画像とテキストのペアを用いて学習していることから, 撮像環境による細かな画角や照明の違いに対しては頑健であることを期待できる. 実際に WinCLIP と AnomalyCLIP のピクセルレベルの精度は, 他手法より Gogan での精度低下の度合いが抑えられており, 期待した効果が得られた. 逆に画像レベルの F1 と AP では MVTec AD に対して Gogan でスコアが大きく低下している. この理由として, MVTec のカテゴリに含まれる”bottle”, ”cable”, “carpet”などと比べて, Gogan で対象とする”concrete block”は CLIP の事前学習での出現頻度が低いと想定され, それに起因して CLIP のイメージエンコーダーで正常ブロックと異常ブロックを判別できる埋め込みベクトルが得られていない可能性があると推察した.

5. おわりに

本研究では, 撮像環境が統一できない異常検知タスクを想定し, 屋外のコンクリートブロックのデータセット(Gogan データセット)を作成した. 撮像環境が統一された MVTec AD とされていない Gogan データセットに対して, 既存の異常検知手法を同じ条件で適用し, データセットによる異常検知性能の比較実験を行った. 実験結果から以下の知見と課題が得られた.

- 撮像環境が統一できない条件では異常検知精度の評価結果が大きく変動することから, 撮像環境が統一されていないベンチマークデータの必要性が示唆される
- 撮像環境が統一できない場合, 正常範囲の過検出が精度低下を招いており, 過検出の抑制が課題となる
- CLIP を特徴量抽出器として使うことで撮像環境の違いに頑健となる可能性があるが, 学習頻度の高くないオブジェクトでの異常検知精度は課題が残る

参考文献

[Bergmann 2019] P. Bergmann, M. Fauser, D. Sattlegger and C. Steger: MVTec AD -- A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection, CVPR, 2019.

[Zou 2022] Y. Zou, J. Jeong, L. Pemula, D. Zhang and O. Dabeer: SPot-the-Difference Self-Supervised Pre-training for Anomaly Detection and Segmentation, ECCV, 2022.

[Roth 2022] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler: Towards total recall in industrial anomaly detection, CVPR, 2022.

[Batzner 2024] K. Batzner, L. Heckler and R. König: EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies, WACV, 2024.

[Zhou 2024] Y. Zhou, X. Xu, J. Song, F. Shen and H. T. Shen: MSFlow: Multi-Scale Normalizing Flows for Unsupervised Anomaly Detection, TNNLS, 2024.

[Jeong 2023] J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran and O. Dabeer: WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation, CVPR, 2023.

[Zhou 2024] Q. Zhou, G. Pang, Y. Tian, S. He and J. Chen: AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection, ICLR, 2024.

[Li 2024] X. Li, Z. Huang, F. Xue and Y. Zhou: MuSc: Zero-Shot Industrial Anomaly Classification and Segmentation with Mutual Scoring of the Unlabeled Images, ICLR, 2024.

[Zhang 2023] Z. Zhang, Z. Zhao, X. Zhang, C. Sun and X. Chen: Industrial Anomaly Detection with Domain Shift: A Real-world Dataset and Masked Multi-scale Reconstruction , Computers in Industry, ELSEVIER, 2023.