Ensemble PatchCore と SQ-VAE による多段階異常検知

都築 幸乃[1,a)]　吉田 龍人[1]　大久保 順一[1]　藤井 純一郎[1]
岡野 将大[1]　緒方 陸[1]　山下 隆義[2]

概要

正常データが多様な物体に対して埋め込みベースの異常検知を行う場合，全ての正常特徴量を網羅した一つの埋め込み空間をつくることは難しい．そこで本研究では，異なる埋め込み空間を持つ異常検知手法を段階的に組み合わせた多段階異常検知アルゴリズムを構築する．そして，ベイズ推定を用いて組合せを最適化する．ここでは PatchCore と SQ-VAE を使用し，それぞれ単体で用いた場合よりも提案手法の異常検知精度が向上することを実験的に示す．

1. はじめに

機械学習を活用したインフラ構造物の維持管理作業の効率化・高度化が期待されている．特に，河川護岸は河川の上流から下流まで広範囲にわたって護岸ブロックが積み重なっており，目視での異常検知は作業者に多大な負担がかかる．そのため，河川護岸の護岸ブロックを対象とした異常検知は重要である．先行研究[1][2]では，護岸の全体画像から河川ブロックを切り出し，ブロック画像ごとに異常検知を行っている．

一方，護岸ブロックの異常は，微小なひび割れからブロックが抜ける大きな変状まで様々である．また，使用されるブロックの形状や積み方によって多くの護岸タイプが存在する．さらに屋外環境のため，雨や太陽光などの影響によって同じ条件下でデータを取得することが難しい（Appendix A）．これらのことから，河川護岸における異常検知では，様々なサイズの変状を検出でき，且つブロックの種類や天候の変化等にロバストな手法を確立することが求められる．

PatchCore[3]などの埋め込みベースのアルゴリズムは異常検知の主要なアプローチの一つであり，異常検知タスクで高い性能を達成している．埋め込みベースの手法は，訓練用の正常データのみを用いて正常な特徴量空間を定義する．そして，入力した特徴量が正常な特徴量空間から離れている場合，異常として検出できる．しかしながら，正常データが多様である護岸ブロック画像を 1 つの正常な特徴量空間に定義することは困難である．そこで本研究では，異なる埋め込み空間を持つ異常検知手法を組み合わせることで正常データの多様性に対応することを考える．

本研究では，多様性がある正常データに適用可能な異常検知手法を確立することを目的として，PatchCore と SQ-VAE[4]を段階的に組み合わせた多段階異常検知アルゴリズムを構築する．そして，ベイズ推定を用いて組合せを最適化する．ここで，従来の VAE は入力画像と再構成画像の相違をもとに異常度合いを算出する．本研究では，SQ-VAE のコードブックを正常特徴量の埋め込み空間として利用した異常検知手法も提案する．

組合せ最適化には教師データとして異常データが必要となるが，提案手法を現場導入するシナリオとして，定期点検によって異常データを経年的に蓄積し，パラメータの更新を繰り返して精度を改善していくことを想定している．

提案手法の有効性を評価するために，本研究では PatchCore と SQ-VAE それぞれ単体の手法と提案手法の異常検知精度の比較を行う．また，提案手法の汎用性を評価するため，2 種類の護岸ブロック画像に対して実験を行う．

2. 異常検知手法

2.1 PatchCore

PatchCore は，正常データの近傍には正常データがあり，異常データの近傍には正常データがないという仮定に基づく異常検知手法である．具体的には，入力した特徴量に対してメモリバンクに保存された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．

2.2 SQ-VAE

SQ-VAE は，ベクトル量子化を用いて入力画像を潜在変数に変換する VAE である（**図-1**）．ベクトル量子化とは，

[1] 八千代エンジニヤリング株式会社
[2] 中部大学
[a)] yk-tsuzuki@yachiyo-eng.co.jp

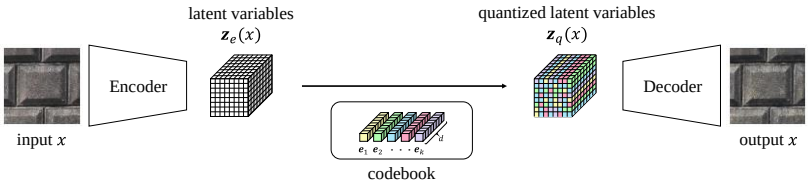
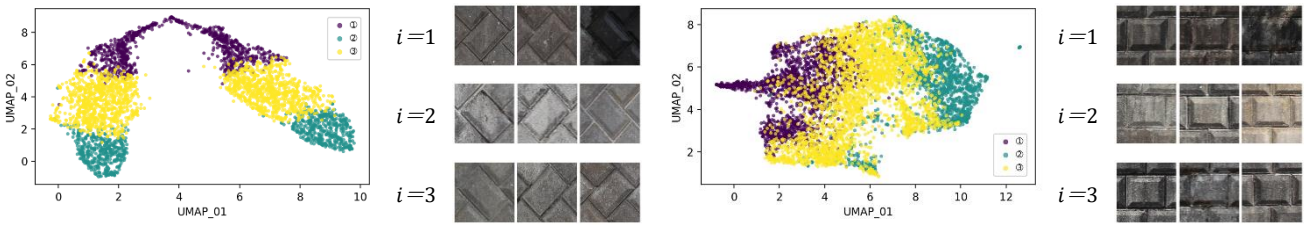


図-1　SQ-VAE 概要



(a)　谷積みブロック　　(b)　布積みブロック

図-2　UMAP と k-means 法によるデータサンプリング

あるベクトルをコードブックと呼ばれるベクトル集合内の埋め込みベクトルに変換する処理である．SQ-VAE では，エンコーダが出力する特徴ベクトル$\boldsymbol{z}_{ei}$とコードブックの埋め込みベクトル$\boldsymbol{e}$の距離が近づくように学習され，式(1)のように，各特徴ベクトルは距離が最も近いk番目の埋め込みベクトル$\boldsymbol{e}_k$に置換される．

$$\boldsymbol{z}_{qi}(x) = \boldsymbol{e}_k,$$
$$where \;\; k = \underset{j}{\operatorname{argmin}} \left\| \boldsymbol{z}_{ei}(x) - \boldsymbol{e}_j \right\|_2 \qquad (1)$$

ここで，$\boldsymbol{z}_e$は量子化前潜在変数，$\boldsymbol{z}_q$は量子化後潜在変数である．

モデルが正常データのみを学習してコードブックを構築した場合，異常領域の特徴ベクトルと置換対象となる埋め込みベクトルの距離は，正常領域の特徴ベクトルとの距離よりも遠くなると仮定する．この仮定に基づき，式(2)のように置換前の特徴ベクトルと置換後の埋め込みベクトルの距離を異常スコアとする．

$$anomaly \;\; score_{SQ-VAE} = \left\| \boldsymbol{z}_{ei}(x) - \boldsymbol{e}_k \right\|_2 \qquad (2)$$

異常スコアは特徴ベクトルごとに得られるため，PatchCore と同様にパッチレベルでも異常を判定することができる．

3.　予備実験

本章では予備実験として，護岸ブロック画像に対する PatchCore と SQ-VAE それぞれの異常検知性能を調査する．

3.1　データセット

2 種類の護岸ブロック画像を使用する．訓練データ用として，谷積みブロック画像を 3,463 枚，布積みブロック画像を 6,063 枚用いる．訓練データで定義される正常な特徴量空間によって異常検知精度は変化する．本研究では，色合いが異なるデータセットと精度の関係性を調査するため，**表-1** に示すように，2 種類のサンプリング方法を用いて 4 種類のデータセットを作成した．まず，**図-2** に示すように，画素値を入力として UMAP による次元削減を行い，さらに k-means 法を用いて全体を 3 つのグループに分割した．それぞれをデータセット 1～3 とし，各データセットは主に色合いが似た護岸ブロック画像で構成されている．次に，谷積みブロック画像は 3,463 枚を全て使用し，布積みブロック画像は 3,000 枚をランダムにサンプリングした．これをデータセット 4 とする．

表-1　データセット内訳

	谷積み	布積み	サンプリング方法
1	830	1,761	k-means 法による分割
2	904	1,488	k-means 法による分割
3	1,729	2,814	k-means 法による分割
4	3,463	3,000	ランダムサンプリング
全体	3,463	6,063	-

テストデータは，性能評価用と，後述する本実験で用いるパラメータ最適化用の 2 種類のデータセットを作成した．それぞれ，谷積みブロック画像は 118 枚，布積みブロック画像は 124 枚使用する．谷積みブロックの異常は 3 クラス（ひび割れ・目地開き，剥離・欠損・抜け，エフロレッセンス），布積みブロックの異常は 4 クラス（ひび割れ・目地開き，剥離・欠損・抜け，エフロレッセンス，溶出物）あり，異常画像には各クラス 6 枚ずつ，正常画像にはどちらも 100 枚用いる．

3.2　訓練と性能評価指標

谷積みブロックと布積みブロックで，データセット 1～4 をそれぞれ訓練した PatchCore（$i = 1$~4）と SQ-VAE（$i = 1$~4）の合計 16 個の検出器を作成した．

異常マップは，PatchCore と SQ-VAE ともに，各パッチの異常スコアを入力画像に対応する位置に並べ，入力画像の

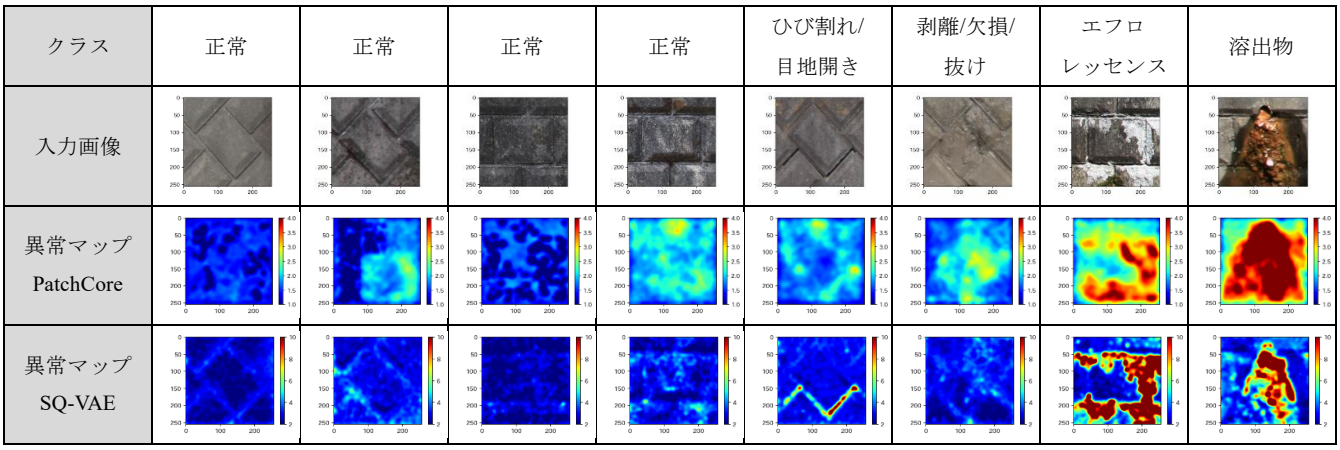


図-3　PatchCore と SQ-VAE の異常検知結果例

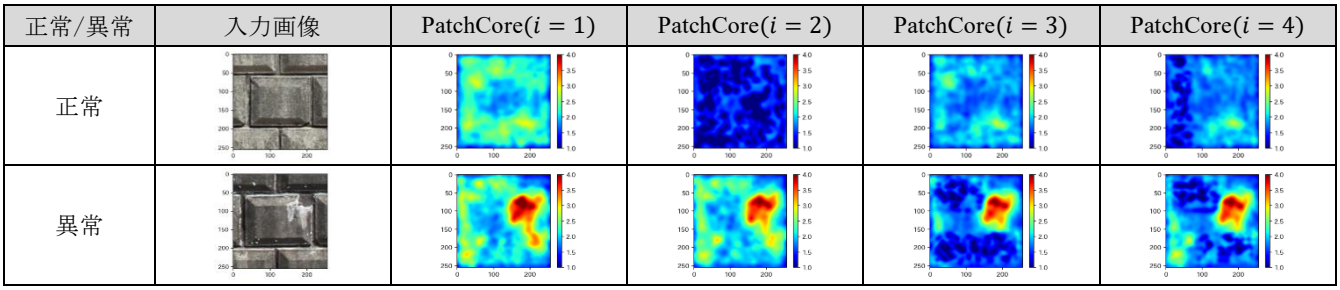


図-4　同じ入力画像に対する PatchCore（$i = 1$~4）の比較

サイズ（256×256）にリサイズした後，ガウシアンフィルタを適用する．ここでは異常マップ全体の最大値を画像レベルの異常スコアとし，AUROC によって異常検知精度を評価する．

3.3 各異常検知手法の特性

性能評価用のテストデータに対する PatchCore と SQ-VAE の異常検知結果例を**図-3** に示す．ここでは AUROC が最も高い検出器の結果のみを示している（谷積み：PatchCore（$i = 4$）と SQ-VAE（$i = 2$），布積み：PatchCore（$i = 4$）と SQ-VAE（$i = 1$））．

PatchCore は実際の異常サイズより広域の異常スコアが高くなる傾向が見られるが，ほぼ全ての異常に高い異常度を示すことを確認した．ただし，影などを含む正常領域を誤検知する場合が見られた．

SQ-VAE は，エフロレッセンスや溶出物などの色の変化が明確な異常に対して敏感に反応する．そのため剥離・欠損のようなブロックの色の変化が小さい異常は検出が難しい．一方で，PatchCore で誤検知されやすい影などを含む正常領域は正しく正常として判定できている．SQ-VAE の異常検知結果の考察については Appendix B で詳細を述べる．

3.4 データセットと精度の関係

同じ入力画像に対する PatchCore（$i = 1$~4）の異常検知結果を**図-4** に示す．これより，異常領域は全ての PatchCore で高い異常スコアを示したが，正常領域は検出器によって異常スコアが大きく異なった．このことから，護岸ブロック画像においては，単一の PatchCore よりも複数の小規模データセットを用いた PatchCore によるアンサンブルな異常検知（以下 ensemble PatchCore とする）が有効な手法であるといえる．

4. 多段階異常検知

本研究では，**図-5** に示すような ensemble PatchCore と SQ-VAE を用いた多段階異常検知アルゴリズムを構築する．

4.1 Step1: ensemble PatchCore

まず，画素値に基づいてグループ分けした小規模データセットで訓練した ensemble PatchCore を用いて複数のメモリバンクを作成する．そして，入力画像に対してそれぞれの PatchCore を用いて異常マップ$P_j(j = 1, \ldots, i)$を出力する．予備実験と同様に異常マップ全体の最大値を画像レベルの異常スコアとし，全ての異常スコアが閾値th_p以下の画像を全て正常（$P = 0$）と判定する．本研究では，予備実験で作成した PatchCore（$i = 1$~3）を使用する．

4.2 Step2: mask with SQ-VAE

予備実験と同様の方法で出力した SQ-VAE の異常マップに閾値th_mを適用して 2 値マスクMを生成する．さらに，マスクMにおいて閾値th_n未満の領域をノイズとして除去する．本研究で使用する SQ-VAE は，予備実験のベストモデル（谷積み：$i = 2$，布積み：$i = 1$）である．

4.3 Step3: final output

Step1 で出力した異常マップを合計し，新しい異常マッ

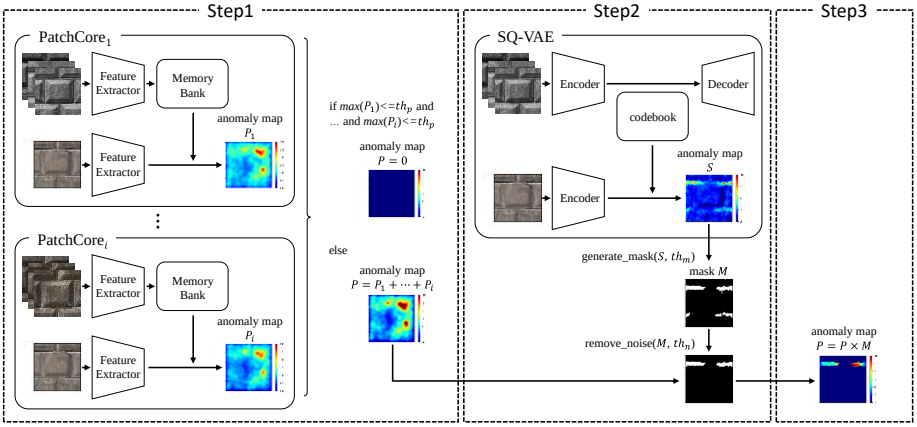


図-5　多段階異常検知の流れ

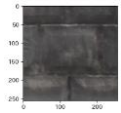
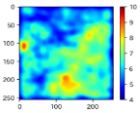
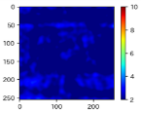
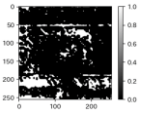
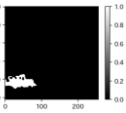
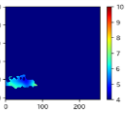
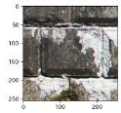
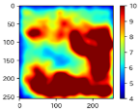
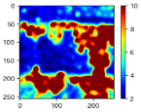
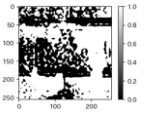
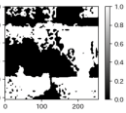
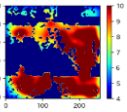
正常/異常	入力画像	異常マップ Step1	異常マップ SQ-VAE	マスク	マスク (ノイズ除去済)	異常マップ Step3
正常						
異常						

図-6　正常画像と異常画像に対する各異常マップとマスク例

表-2　性能評価結果（AUROC）

異常検知手法	谷積み	布積み
PatchCore（best model）	0.9566	0.9575
SQ-VAE（best model）	0.9272	0.8675
step1（ensemble PatchCore）	0.9583	0.9704
step2（mask with SQ-VAE）	0.9750	0.9529
step3（step1+step2）	**0.9833**	**0.9804**

プ$P = P_1 + \cdots + P_i$を生成する．この異常マップPにマスクMを乗算したものを最終的な異常マップ$P = P \times M$とする．

4.4　パラメータ最適化

ベイズ推定を用いて各パラメータth_p，th_m，th_nの最適化を行う．本研究では，最適化には optuna[5]を用いる．最適化の目的関数は$1 - \mathrm{AUROC}$とし，最適化用のテストデータを用いて試行回数 1000 回で最適化を行った．

5.　実験

性能評価用のテストデータに対する PatchCore と SQ-VAE，提案手法の AUROC を**表-2** に示す．これより，谷積みブロックと布積みブロックどちらにおいても提案手法が最も高い精度を達成している．また，単体の PatchCore よりも ensemble PatchCore を用いた Step1 の精度が高く，さらに step1 よりも SQ-VAE を組み合わせた Step3 の精度が高くなっていることから，多段階による異常検知手法の有効性を確認できる．

正常画像と異常画像に対する各異常マップとマスクの例を**図-6** に示す．正常画像では，Step1 で異常スコアが高い正常領域が Step3 でマスクされており，異常マップ全体の最大値を画像レベルの異常スコアとしたとき，異常度が低い結果となっている．異常画像では，マスクされた後も異常マップに異常領域が残っており，異常度が高い結果となっている．このように提案手法では，ensemble PatchCore が全ての異常を検知できているという前提のもと，ensemble PatchCore が過剰に異常と判定した領域を SQ-VAE の異常マップを利用することによって抑制している．

6.　おわりに

本研究では，正常データが多様な場合でも頑健な異常検知を実現することを目的として，PatchCore と SQ-VAE を組み合わせた多段階異常検知アルゴリズムを構築した．そして，ベイズ推定を用いて組合せを最適化した．予備実験にて，護岸ブロック画像に対する PatchCore と SQ-VAE の異常検知特性の違いを調査し，さらに ensemble PatchCore の有効性を示した．次に，提案手法が単体の手法よりも精度が高いことを確認した．また，2 種類の護岸ブロック画像に対して同様の結果が得られ，提案手法に汎用性があることを示した．

今後は，最適化に用いる異常データのクラスやデータ数などのデータ基準の明確化に取り組む．

References

[1] Y. Tsuzuki, R. Yoshida, J. Okubo, J. Fujii, and T. Yamashita, Anomaly detection of revetment by unsupervised deep learning using variational auto-encoder, JSCE, 2022.
[2] Y. Tsuzuki, R. Yoshida, J. Okubo, and J. Fujii, Comparison study of continuous and discrete latent variable models, JSAI, 2023.
[3] K. Poth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, Towards total recall in industrial anomaly detection, CVPR, 2022.
[4] Y. Takida, T. Shibuya, W. Liao, C.-H. Lai, J. Ohmura, T. Uesaka, N. Murata, S. Takahashi, T. Kumakura, and Y. Mitsyfuji, SQ-VAE: Variational bayes on discrete representation with selfannealed stochastic quantization, ICML, 2022.
[5] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: a next-generation hyperparameter optimization framework, KDD, 2019.