異常検知のための SQ-VAE コードブックへのアプローチ

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

概要

本研究では，河川護岸の様々な形状のブロックに対して異常検知を行うために，SQ-VAE のコードブックを異常検知に利用することを検討する．解析の結果，SQ-VAE のコードブックには異常な特徴に対応するコードが存在し，そのコードが割り当てられたパッチのエンコードベクトルは正常と異常に分離できる可能性があることを発見した．さらに本研究では，異常検知に有用なコードを効率的に使用するための有効なネットワーク構造とコードブックサイズについて調査した．

1. はじめに

河川護岸を点検する際，ICT 技術等の積極的な活用による現場作業の効率化・高度化に取り組むことが望まれている．これに対し，先行研究[1]では DIP-VAE[2]を用いた異常検知手法を提案している．これは，一つの護岸ブロックを均一な部品として捉え，ブロック毎に異常検知を行う手法である．この手法は再構成誤差に基づく評価を行うため，モデルが再構成困難な護岸ブロックに対しては異常検知性能が著しく低下することが問題である．

従来，VAE は他の生成モデルに比べて低解像度な画像が生成されるため，再構成能力の低さが指摘されていた[3]．その中で，VQ-VAE[4]はコードブックを用いて量子化した離散的な潜在変数を使用することにより，VAE ベースのアプローチで高精度な画像生成を可能にした．この VQ-VAE をさらに改良したモデルが SQ-VAE[5]である．SQ-VAE は，学習の進行に伴って確率的量子化から決定論的量子化へと徐々に変更させることで，コードブック崩壊[6]が発生しにくい適切なコードブックの学習を促している．

本研究では，様々な護岸ブロックに適用可能な異常検知手法を確立することを目的とし，SQ-VAE の異常検知への利用可能性を調査する．ただし，再構成能力の高い SQ-VAE は学習していない特徴まで再構成できてしまう場合があり，再構成ベースの異常検知を行うことは困難である．そこで本研究では，SQ-VAE のコードブックを活用した異常検知手法について検討する．まず，正常データと異常データに対するコードブックの割り当て傾向を解析し，コードブック上での正常と異常の分離を試みる．次に，解析で得られた異常検知に有用なコードを効率的に使用するために，モデルのネットワーク構造とコードブックサイズを変更し，それらの有効性について調査する．実験で得られた知見の一般性を確認するため，本研究では護岸ブロック画像の他に工業用異常検知データセット MVTec AD[7]のヘーゼルナッツ画像でも調査を行う．

[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

2. SQ-VAE

VQ-VAE では，エンコーダが出力した特徴マップの各ベクトル（以下エンコードベクトルとする）をコードブックの埋め込みベクトル（以下コードベクトルとする）の中から最も距離が近いもので置換する決定論的量子化が行われる．一方 SQ-VAE では，学習の初期段階では確率的量子化が行われ，学習が進むにつれて徐々に VQ-VAE と同じ決定論的量子化へと収束する．この過程をセルフアニーリングと呼ぶ．SQ-VAE への入力が画像や音声といった連続値の場合，エンコードベクトルはコードベクトルをガウス分布に従って逆量子化したものと仮定される．この仮定に基づいて確率的量子化過程が導出され，学習によって分散が 0 になると確率 1 でエンコードベクトルに最も近いコードベクトルが選択される．

コードブックの各コードがそれぞれ異なるガウス分布を持ち，入力画像中の正常な特徴と異常な特徴の分布が異なるとすると，その特徴に割り当てられるコードも異なると推測する．この考えを裏付けるため，次章では SQ-VAE のコードブックを解析する．

3. コードブック解析

コードブックを利用する VAE の潜在変数は，入力画像を任意のサイズにパッチ分割し，それにコードブックをマッピングしたものとして見ることができる．実際に潜在変数を可視化した**図-1** のカラーマップは，同じ色のパッチには

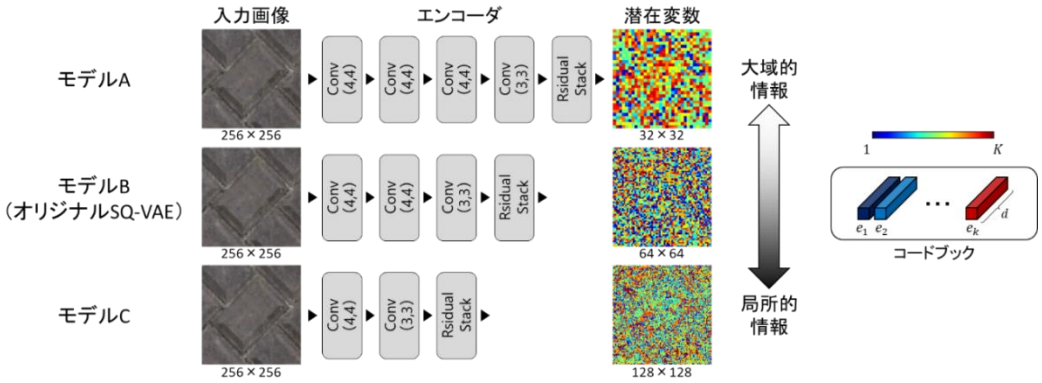


図-1　潜在変数の可視化とネットワーク設計

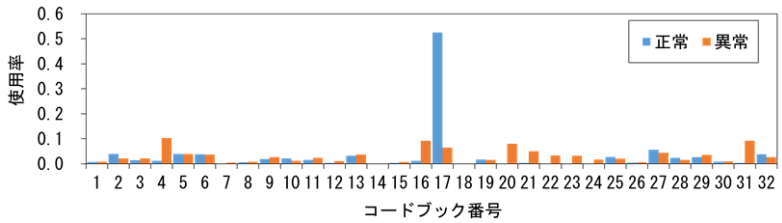


図-2　正常/異常データに対するコードブックの度数分布
（パッチ総数で正規化）

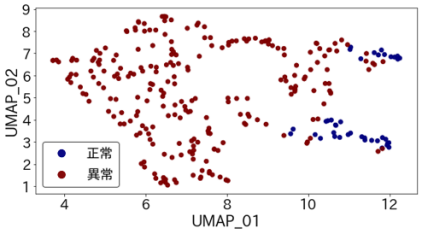


図-5　UMAP による
エンコードベクトルの次元削減

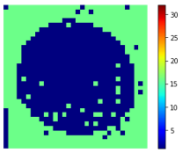
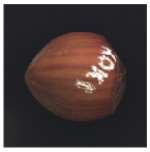
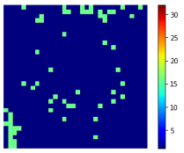
入力画像（正常）	17 番のコード	入力画像（異常）	17 番のコード
			

図-3　正常データで頻出するコードの抽出結果（対象コード以外を 0 に置換）

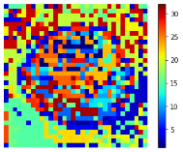
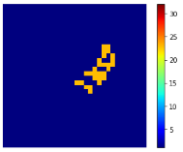
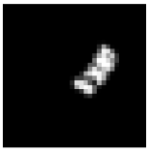
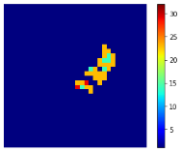
入力画像 （256×256）	潜在変数 （32×32）	24 番のコード	Ground Truth （32×32）	Ground Truth に 対応するコード
				

図-4　異常（print）に対応するコードの抽出結果（対象コード以外を 0 に置換）

同じ番号のコードが割り当てられていることを意味する．ここでは入力画像中の正常な特徴のみを含むパッチを正常パッチ，異常な特徴を含むパッチを異常パッチとする．

3.1　データセット

本研究では，護岸ブロック画像 376 枚とヘーゼルナッツ画像 391 枚を学習データとして使用する．回転, 平行移動, 左右反転を行い，護岸ブロック画像は 6,016 枚，ヘーゼルナッツ画像は 12,121 枚に拡張した．

3.2　解析

ヘーゼルナッツ画像を対象に，1 枚の画像に対するパッチ分割数を 32×32 に設定する．正常データのパッチ（32×32×40＝40,906 個）と異常データのパッチ（32×32×17＝17,408 個）に対するコードブックの度数分布を**図-2** に示す．正常データでは 17 番のコードが頻出しており，このコードは正常画像の背景部分に対応していることが分かった（**図-3**）．一方，異常データで特に頻出しているコードはなかったため，異常データに対する各コードの抽出結果を目視で確認すると，24 番のコードが異常に対応していることを発見した（**図-4**）．24 番のコードは正常データで 39 個，異常データで 300 個のパッチに割り当てられ, 300 個中 245 個は異常パッチに該当し，これはリサイズした Ground Truth（32 × 32）の約 55%を占めていた．さらに，39 個の正常パッチと 245 個の異常パッチのエンコードベクトルを

表-1　各モデルの学習データに対するコードブック使用率

	モデル	サイズ	使用率	サイズ	使用率	サイズ	使用率
護岸ブロック	A	32	**0.9688**	128	0.8594	512	0.8066
	B	32	0.9688	128	0.9531	512	**0.9727**
	C	32	1.0000	128	**0.9766**	512	0.9688
MVTec AD hazelnut	A	32	**0.9688**	128	0.9219	512	0.7227
	B	32	0.9688	128	**0.9844**	512	0.9746
	C	32	0.9063	128	**0.9688**	512	0.8965

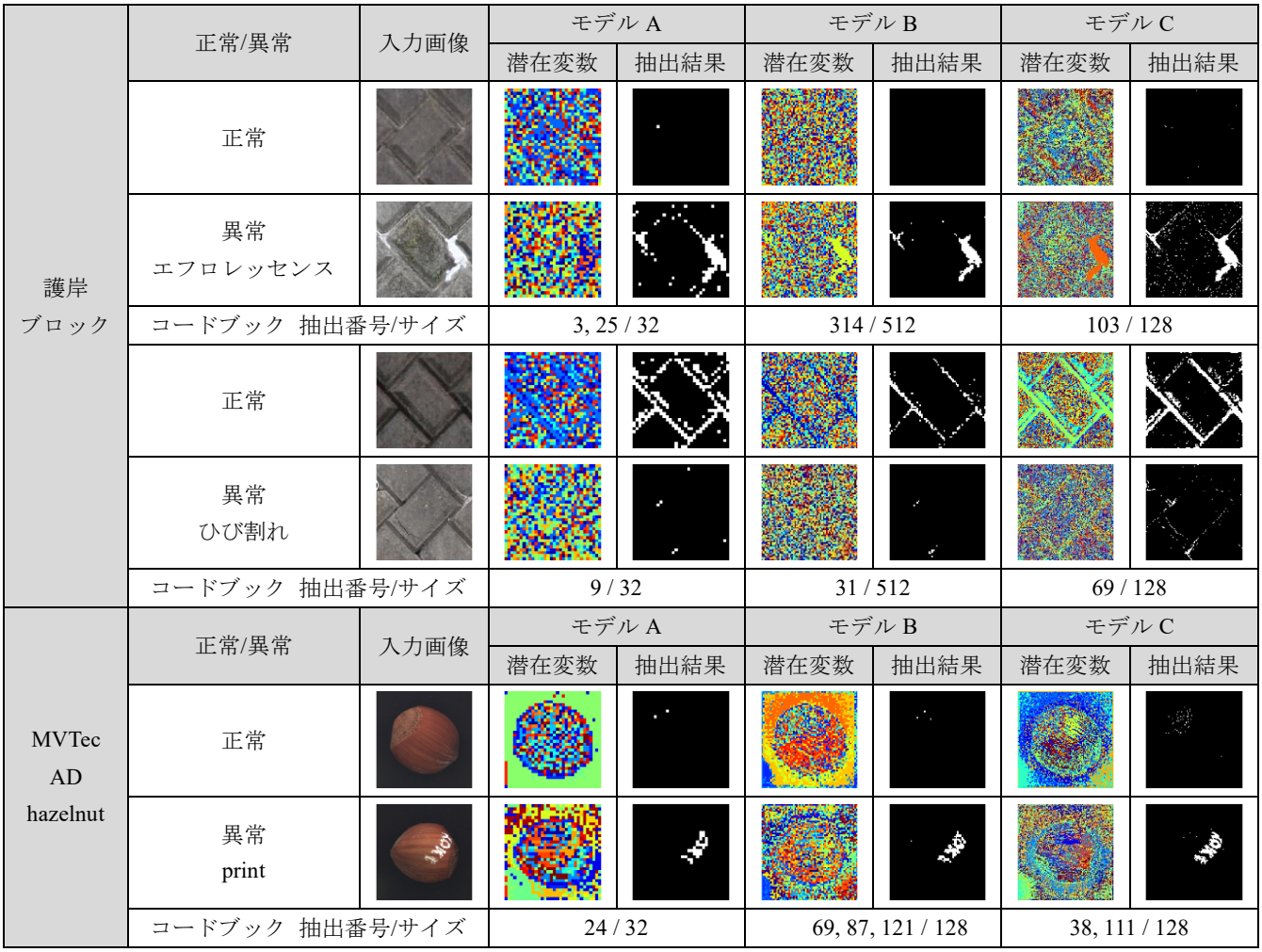


図-6　各ネットワークにおける異常コードの抽出結果

UMAP[8]を用いて次元削減した結果，同一のコードの中で正常と異常のクラスタが分かれる傾向があった（**図-5**）．

以上の結果より，24 番のコードのように異常パッチに占める割合が多いコード（以下異常コードとする）は異常検知に有用であると考える．

4. 異常コード抽出へのアプローチ

入力画像をエンコードする際．畳み込みの回数によって潜在変数に射影される情報は変化する．例えば階層的な潜在変数を導入した VQ-VAE2[9]では，畳み込み回数の違いによって局所的な情報と大域的な情報を持つ潜在変数がある．このことから，潜在変数の持つ情報によって捕捉できる異常は変化すると考える．また，VQ-VAE や SQ-VAE はコードブックサイズを増加させるほど表現能力が高まることが知られており，コードブックサイズがモデルに与える影響は非常に大きい．本章では，異常コードを効率よく異常検知に使用するために，ネットワーク構造とコードブックサイズをそれぞれ変更しながら有効なモデル設定について調査する．具体的には，エンコーダ・デコーダの畳み込み層の数が異なる 3 種類（A : 4 層，B : 3 層，C : 2 層）のネットワークを設計し，3 種類（K＝32，128，512）のコードブックサイズを持つモデルをそれぞれ作成する（**図-1**, **表-1**）．

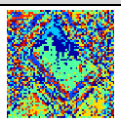
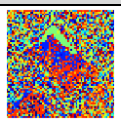
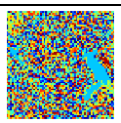
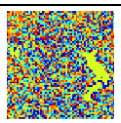
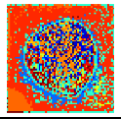
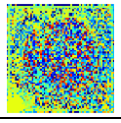
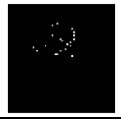
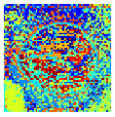
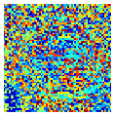
	正常/異常	入力画像	潜在変数	抽出結果	潜在変数	抽出結果
護岸ブロック	正常					
	異常 エフロレッセンス					
	コードブック 抽出番号/サイズ		11 / 32		314 / 512	
	正常/異常	入力画像	潜在変数	抽出結果	潜在変数	抽出結果
MVTec AD hazelnut	正常					
	異常 print					
	コードブック 抽出番号/サイズ		16, 29 / 32		176, 202, 287, 416, 491 / 512	

図-7　コードブックサイズが異なるモデル B における異常コードの抽出結果

4.1 ネットワーク構造の変更

各ネットワークにおいて，護岸ブロックは目視，ヘーゼルナッツは Ground Truth と照らし合わせて異常コードをそれぞれ抽出した（**図-6**）．ここでは，学習データに対するコードブック使用率が高いコードブックサイズのモデルを採用している（**表-1**）．潜在変数が局所的な情報を持つモデル C は，モデル A やモデル B では確認できない細いひび割れや印字の詳細部分を捕捉することができている．一方で，エフロレッセンスや印字の大部分はモデル A でも捕捉することができている．

4.2 コードブックサイズの変更

コードブックサイズが 32 個と 512 個のモデル B において，前節と同様に護岸ブロックとヘーゼルナッツの異常コードをそれぞれ抽出した（**図-7**）．白い沈着物質であるエフロレッセンスに対応する異常コードは，ブロックの色が明るい正常画像のパッチにも多く割り当てられている．しかし，コードブックサイズを 32 個から 512 個に変更した場合，異常画像の結果はほとんど変化しないが，正常画像で異常コードが割り当てられるパッチ数が減少した．このことから，コードブックサイズが大きいほどコードブック上で正常と異常が分離しやすくなると推測する．一方で，ヘーゼルナッツの結果ではコードブックサイズの増加に伴って異常コードの数が 2 個から 5 個に増加した．これより，コードブックサイズが大きくなると同じ種類の異常が複数のコードに細分化されると考える．

5. まとめ

本研究は，様々な護岸ブロックに適用可能な異常検知手法を確立することを目的とし，SQ-VAE のコードブックを異常検知に利用することを検討した．

まず正常データと異常データに対するコードブックの割り当て傾向を解析し，異常な特徴に対応するコードが存在することを発見した．さらに，そのコードが割り当てられたエンコードベクトルが正常と異常に分離できる可能性を示した．異常コードを発見し，それが割り当てられた正常パッチと異常パッチを分離する一連の流れを手法として確立することができれば，SQ-VAE による異常検知が可能になると考える．

次に，異常コードを効率よく使用するために，有効なネットワーク構造とコードブックサイズについて調査した．ネットワーク構造を変更した結果では，潜在変数が局所的な情報を持つほど高い空間解像度で異常を捕捉できることを確認した．そのため，検出対象とする異常の大きさによってネットワーク構造を変更する必要があると考える．また，本研究では畳み込み層の数が異なるモデルを個別に作成したが，階層構造を持つモデルであれば一つのモデルで網羅的に異常を捕捉できる可能性があり，これは今後の検討とする．コードブックサイズを変更した結果では，コードブックサイズが大きいほどコードブック上で正常と異常がより分離される傾向が見られた．ただし，際限なくコードブックサイズを増加させると同じ種類の異常が複数のコードに細分化され，異常検知の際に注目するべきコードの発見が困難になるおそれがある．このように，SQ-VAE のコードブックを異常検知に利用するためには，小さすぎず大きすぎない適切なコードブックサイズを設定する必要があるが，その値は入力データに依存するものと考える．

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] A. Kumar, P. Sattigeri, and A. Balakrishnan, “Variational inference of disentangled latent concepts from unlabeled observations”, ICLR, 2018.
[3] G. Perarnau, J. Van De Weijer, B. Raducanu, and J, M, Alvarez, “Invertible conditional gans for image editing”, NIPS, 2016.
[4] A. van den Oord, O. Vinyals, K. Kavukcuoglu, “Neural discrete representation learning“, NIPS, 2017.
[5] 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.
[6] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Shtskever, “Jukebox: A generative model for music”, arXiv preprint arXiv: 2005.00341, 2020.
[7] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, “MVTec AD – A comprehensive real-world dataset for unsupervised anomaly detection”, CVPR, 2019.
[8] L. Mclnnes, J. Healy, and J. Melville, “UMAP: Uniform manifold approximation and projection for dimension reduction” , arXiv preprint arXiv: 1802.03426, 2018.
[9] A. Razavi, A. van den Oord, and O. Vinyals, “Generating diverse high-fidelity images with VQ-VAE-2”, NIPS, 2019.