VAEを用いた河川護岸の異常検知における 連続潜在変数モデルと離散潜在変数モデルの比較 Comparison study of continuous and discrete latent variable models for anomaly detection of river revetment using VAE

都築 幸乃^{*1} 吉田 龍人^{*1} 大久保 順一^{*1} 藤井 純一郎^{*1} Yukino Tsuzuki Ryuto Yoshida Junichi Okubo Junichiro Fujii

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

An anomaly detection method for river revetment using VAE is being studied. This previous study was conducted for a single block shape, and it is necessary to investigate a method that can be applied to blocks of various shapes in the future. However, it has been pointed out that the reconstruction performance of VAE is low. Hence, there is a possibility that some blocks may be difficult to reconstruct. On the other hand, VQ-VAE, which enables high-resolution image generation by using discrete latent variables, and SQ-VAE, which further improves its performance, have been proposed. In this study, we investigate the characteristics of both the continuous latent variable model (DIP-VAE) used in the previous study and the discrete latent variable model (SQ-VAE) with high reconstruction performance, with the aim of performing anomaly detection on various types of blocks.

1. はじめに

現在の河川護岸は評価要領[国土交通省 2019]に従い,主 に目視点検によって管理されているが,点検者によって評価に バラつきがあることや,技術者数の減少等が問題とされている. 近年では,コンクリート構造物の点検における深層学習を活用 した研究が盛んに行われており,筆者らもその一手法として, VAE (Variational AutoEncoder) [Kingma 2014]を用いた河川護 岸の異常検知手法[都築 2022]を提案している. VAE は,オブ ジェクトの中に共通する特徴が学習されるため,観測される確率 が低い特徴の再構成は困難とされている.この特性を利用し, 一般に,入力データと再構成データの類似度を異常スコアとし て評価することができる.先行研究では,工業製品の不良品検 知等に倣い,河川護岸を構成する一つのブロックを均一な部品 として捉え,ブロック単位での異常検知を試みた.その結果,河 川護岸の点検において,特定の変状種別に依存しない異常検 知技術を確立することができた.

先行研究はブロック形状が単一の場合の検討であり、今後は 実用化に向けて、様々な形状のブロックに適用可能な異常検 知手法を検討する必要がある.しかし先行研究では、学習させ る正常データのバリエーションを増やした(植生や水抜口等の オブジェクトを追加した)場合、それらをうまく再構成することが できず、異常検知性能が低下した.このことから、ブロック形状 が多様になった場合、先行研究手法では再構成および異常検 知が困難なブロックが存在する可能性がある.

従来, VAE は GAN などの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, その再構成性能の低さが問題とさ れていた.その中で, 再構成性能が非常に高い VQ-VAE (Vector Quantised Variational AutoEncoder)[Oord 2017]が提案 された. VQ-VAEは、オリジナルの VAE が正規分布に従う連続 的な潜在変数を用いるのに対し、コードブックを用いた離散的

連絡先:都築幸乃, 八千代エンジニヤリング株式会社, 東京都 台東区浅草橋 5-20-8 CS タワー3F 技術創発研究所, TEL: 03-5822-7032, Mail: yk-tsuzuki@yachiyo-eng.co.jp

な潜在変数を使用することで,高解像度な画像生成を可能とした.それぞれの概要を図-1 に示す.さらに, VQ-VAE のフレームワークを改良して再構成性能をより向上させたモデルが SQ-VAE (Stochastically Quantized Variational AutoEncoder) [Takida 2022]である. VQ-VAE では潜在変数を確定的に求めるが, SQ-VAE では確率的に潜在変数を選択することにより,コードブック崩壊(すべてのエンコーディングが単一または少数の埋め込み ベクトルにマッピングされ,コードブック内の他の埋め込みベクトルが使用されない現象) [Dhariwal 2020]を起こさない適切なコードブックが学習されると見られている.

本研究では、様々なブロックで異常検知を行うことを目的とし、 先行研究で用いた DIP-VAE (Disentangled Inferred Prior Variational AutoEncoder) [Kumar 2018]と SQ-VAE のそれぞれ のモデルの特性について調査する.実験に使用するブロック画 像は、四角錐形の石材(間知石)を水平に積んだ布積みブロッ クと、斜めに積んだ谷積みブロックの2種類である.



(b) コードブックを用いる VAE (離散潜在変数モデル)図-1 VAE の概要

2. VAE モデル

2.1 VAE

VAE は、入力データを低次元の潜在変数に圧縮するエンコ ーダと、潜在変数から入力データを再構成するデコーダで構成 される.まず潜在変数zをサンプリングし、zをデコーダ $P_{\theta}(x|z)$ に 入力してデータが生成される.サンプルxごとの負の ELBO は次 式で表される.

$$\mathcal{L}_{VAE} = \mathbb{E}_{z \sim Q_{\phi}(z|x)} [-log P_{\theta}(x|z)] + KL [Q_{\phi}(z|x) || P(z)]$$
(1)

ここで、 $Q_{\phi}(z|x)$ は事前分布P(z)の近似事後分布であり、 $\theta \geq \phi$ は学習可能なパラメータである. ELBO を解析的に計算するに は近似事後分布 $Q_{\phi}(z|x)$ を条件付きガウス分布でモデル化する のが一般的であり、オリジナルの VAE ではターゲットとなる事前 分布P(z)を標準正規分布 $\mathcal{N}(0,I)$ と仮定している. VAE は、 ELBO の最大化、つまり式(1)を最小化することによって潜在変 数モデルを学習することを目的とする.

2.2 DIP-VAE

VAE の派生モデルとして, disentangle な潜在変数を取得す ることを目的としたモデルが提案されている. Disentangle な潜在 変数とは、ある一次元の潜在変数がオブジェクトの位置や大き さ、色などの特定の表現に対応したものを指す. DIP-VAE では、 潜在変数の disentangle 化を促すために、 $Q_{\phi}(z) \ge P(z) \circ 2 \circ 0$ 分布のモーメントを一致させることが提案されている. 具体的に は、 $Q_{\phi}(z) \ge \mathcal{N}(0, I)$ の共分散を一致させ、次式で表される目的 関数 (Disentangled Inferred Prior objective) が提案された.

$$\mathcal{L}_{DIP-VAE} = \mathcal{L}_{VAE} + \lambda_{od} \sum_{i \neq j} \left[Cov_{Q_{\phi}(z)}[z] \right]_{ij}^{2} + \lambda_{d} \sum_{i} \left(\left[Cov_{Q_{\phi}(z)}[z] \right]_{ii} - 1 \right)^{2}$$

$$(2)$$

ここで、 $Cov_{Q_{\phi}(z)}[z]$ は $Q_{\phi}(z)$ の共分散、 λ は目的関数における 正則化項の寄与を制御するパラメータである.2 つの分布の共 分散を一致させることは、事前分布P(z)が標準正規分布 $\mathcal{N}(0,I)$ の場合、 $Q_{\phi}(z)$ の次元を無相関化することに等しい.

2.3 VQ-VAE

VAE が確率的なエンコーダ・デコーダで構成されるのに対し, VQ-VAE は、確定的なエンコーダ・デコーダと学習可能なコー ドブックから構成される. コードブックは一様分布を仮定したK個 の離散的なD次元ベクトル(e_1, e_2, \dots, e_k)で表される. まず入力 データxから特徴マップ $z_e(x)$ (H×W×D)をエンコードし、次に 特徴マップ $z_e(x)$ の各ベクトル(エンコードベクトル)について、コ ードブックの中から一番近い埋め込みベクトルを求める. このe を量子化後の潜在変数 $z_q(x)$ における $z_e(x)$ に対応する位置の ベクトルとする(式(3)). 最後に、すべてのエンコードベクトルを 埋め込みベクトルで置き換えた潜在変数 $z_q(x)$ をデコーダに入 力し、データを生成する.

$$z_q(x) = e_k,$$

where $k = argmin_j ||z_e(x) - e_j||_2$ (3)

VQ-VAE はコードブック崩壊が発生しやすく,そのためコード ブックサイズ(埋め込みベクトルの数)を増やしてもモデルの表 現力が頭打ちになることが問題とされている.

2.4 SQ-VAE

前節で述べた通り、VQ-VAE はエンコードベクトルと距離が 最も近い埋め込みベクトルを確定的に選択する.一方 SO-VAE では,学習初期は確率的に埋め込みベクトルが選択され,学習 が進むにつれて特定の埋め込みベクトルが選ばれる確率が1 に近づき、最終的に距離が最も近い埋め込みベクトルがほとん どの場合に選択されるようになる. つまり, 学習の初期段階では 確率的な量子化が行われ、学習によって徐々に VQ-VAE と同 じ確定的な量子化へと収束する.これをセルフアニーリングと呼 ぶ. エンコードされた量子化前の潜在変数Zは,量子化後の潜 在変数 Z_q を逆量子化した際,その要素 $z_{q,i}$ を平均,学習可能な パラメータΣのを分散としたガウス分布に従うと仮定する $(p_{\varphi}(z_i|Z_q) = N(z_{q,i}, \Sigma_{\varphi}))$. この仮定に基づき確率的量子化過 程 $\widehat{P}_{arphi}ig[Z_q|Zig]$ が導出され,学習によって Σ_{arphi} を 0 に近づけることで 確率 1 でZに最も近いZaが選択される. SQ-VAE では, デコー ダの出力xもガウス分布に従って確率的に計算される $(p_{\theta}(x|Z_{a}) = N(f_{\theta}(Z_{a}), \sigma^{2}I))$. このとき、ガウス分布の平均は ニューラルネットの出力 $f_{\theta}(Z_q)$,分散は $\sigma^2 I$ である. σ^2 は学習可 能なパラメータであり、学習の進行に伴って 0 に近づく. 元論文 では, σ^2 が0に近づけば Σ_{φ} も0に近づくことが証明されている. SQ-VAEの目的関数は以下で表される.

$$\mathcal{L}_{SQ-VAE} = \mathbb{E}_{q_{\omega}(Z|X)\hat{P}_{\varphi}(Z_q|Z)} \left[\frac{1}{2\sigma^2} \| x - f_{\theta}(Z_q) \|_2^2 + \mathcal{R}_{\varphi}^{\mathcal{N}}(Z, Z_q) \right] - \mathbb{E}_{q_{\omega}(Z|X)} H\left(\hat{P}_{\varphi}(Z_q|Z) \right) + \frac{D}{2} \log \sigma^2 + const.$$

$$(4)$$

ここで, $H(\hat{P}_{\varphi})$ は \hat{P}_{φ} のエントロピー, $\mathcal{R}_{\varphi}^{\mathcal{N}}(Z, Z_q)$ は Σ_{φ} の定義に応じた正則化項を指す. ここでは, $\Sigma_{\varphi} = \sigma_{\varphi}^{2}I$ とする.

3. 実験

3.1 データセットとモデル

本研究では、布積みと谷積みの2種類のブロック画像を使用 する.回転やシフト、左右反転のData Augmentationを行い、学 習データとして、布積みブロックの正常画像を13,300枚(元画 像950枚)、谷積みブロックの正常画像を13,920枚(元画像879 枚)用意した.テストデータは、布積みと谷積みそれぞれに対し て正常/異常画像を各100枚用意した.異常画像に含まれる変 状は、表面剥離、目地開き、ひび割れ、エフロレッセンス、溶出 物、欠損である.

本研究では、布積みブロック画像と谷積みブロック画像をそれぞれ学習した DIP-VAEと SQ-VAE の合計 4 つのモデルを作成した.計算条件は表-1 の通りである.

表-1 学習条件

	DIP-VAE SQ-VAE	
入力サイズ	(256, 256, 3)	(256, 256, 3)
エポック数	300	30
潜在変数次元数	128	64
コードブックサイズ	-	512

2N6-GS-10-03

ブロック 種類	正常/ 異常	異常内容	入力画像	再構成画像 DIP-VAE	再構成画像 SQ-VAE	SQ-VAE の 潜在変数
布積み ブロック	正常	-				
	異常	エフロレッセンス				
		表面剥離				
		ひび割れ・目地開き		1		
谷積み ブロック	正常	-	X	$\langle \rangle$		
	異常	エフロレッセンス				
		表面剥離				
		ひび割れ・目地開き				

The 37th Annual Conference of the Japanese Society for Artificial Intelligence, 2023

図-2 各ブロックの正常/異常画像に対する DIP-VAE と SQ-VAE の再構成結果と SQ-VAE の潜在変数の可視化結果

3.2 再構成と異常検知結果

各モデルの再構成結果を図-2 に示す. DIP-VAE は布積み ブロックに対しては入力画像に似た画像を再構成できているが, 谷積みブロックに対しては入力画像と乖離した画像を生成して いる. 一方 SQ-VAE は, 布積みブロックも谷積みブロックも高精 度な再構成ができている. しかし, 正常領域だけでなく異常領 域も再構成できてしまっている.

入力と再構成の MSE を異常スコアとし、布積みブロックと谷 積みブロックそれぞれに対する各モデルの AUROC を求めた (表-2). DIP-VAE は布積みブロックに対する異常検知性能は 高いが、谷積みブロックは再構成が困難であったため、異常検 知性能も低い結果となった. SQ-VAE は正常領域も異常領域も 再構成してしまうため、AUROC の値は低く、再構成誤差による 評価指標は適用できないことが分かった.

3.3 コードブックの分析

既往研究[Marimont 2021]では、VQ-VAE が十分に大きな潜 在空間をもつ場合, 異常な画像を入力しても再構成が可能であ ることが報告されており, 前節より, SQ-VAE でも同様の傾向が 見られることを確認した. そこで本研究では、SQ-VAE のコード ブックを利用した異常検知の可能性について調査する.

図-2 に示すように、コードブックを使用する VAE の潜在変数 は、入力画像のパッチにコードブックを符号化したものとして見 ることができる. 図では 512 個のコードブックを使用しており、カ ラーマップの同じ色のパッチには同じコードが割り当てられてい ることを意味する. 潜在変数の可視化結果より、SQ-VAE は異 常画像を入力しても再構成ができてしまうが、異常領域には正 常領域とは異なるコードが割り当てられてられる傾向があること が確認できた. 特に、布積みブロックの目地開きや谷積みブロ ックのエフロレッセンス、表面剥離が顕著な結果である. さらに、 コードブックサイズを 512 個から 8 個に減少させると、モデルの 表現能力は劣化するが、正常領域と異常領域に割り当てられる コードの違いがより明確になった(図-3).

4. おわりに

本研究は、様々な形状の護岸ブロックで異常検知を行うことを目的とし、連続潜在変数モデル(DIP-VAE)と離散潜在変数 モデル(SQ-VAE)の特性を調査した.DIP-VAE は正常な特徴 のみを再構成するため再構成誤差に基づく異常検知が可能で あるが、再構成が困難な対象の場合、異常検知性能が低い、 一方、高い再構成性能をもつ SQ-VAE は、DIP-VAE では再構 成が困難だったものも再構成できるが、正常領域だけでなく異 常領域も再構成できてしまうため、再構成ベースの異常検知に は適用できないモデルである.そこで SQ-VAE のコードブックに 着目したところ、潜在空間上で正常領域と異常領域が分離でき る可能性が示唆された、今後は、SQ-VAE のコードブックから異 常情報を抽出する手法について検討する.

参考文献

- [国土交通省 2019] 国土交通省 水管理・国土保全局 河川環境 課: 堤防等河川管理施設及び河道の点検・評価要領, 2019.
- [Kingma 2014] Kingma, P. D. and Welling, M.: Auto-Encoding Varia-tional Bayes, ICLR, 2014.
- [都築 2022] 都築幸乃,吉田龍人,大久保順一,藤井純一郎,山 下隆義: VAE を用いたブロック単位での護岸の異常検知, 土木学会 第 25 回応用力学シンポジウム講演概要集, 2022.

表-2 異常検知性能評価結果 (AUROC)

	DIP-VAE	SQ-VAE
布積みモデル	0.9821	0.5458
谷積みモデル	0. 5281	0.5115



SQ-VAE の再構成画像と潜在変数の可視化結果の比較

- [Oord 2017] Oord, van den. A., Vinyals, O. and Kavukcuoglu, K.: Neural discrete representation learning, NIPS, 2017.
- [Takida 2022] Takida, Y., Shibuya, T., Liao, W., Lai, C., Ohmur a, J., Uesaka, T., Murata, N., Takahashi, S., Kumakura, T. an d Mitsufuji, Y.: SQ-VAE: Variational Bayes on Discrete Rep resentation with Selfannealed Stochastic Quantization, ICML, 2020.
- [Dhariwal 2020] Dhariwal, P., Jun, H., Payne, C., Kim, W. J., Ra dford, A. and Sutskever, I.: Jukebox: A Generative Model for Music, arXiv preprint arXiv: 2005.00341, 2020.
- [Kumar 2018] Kumar, A., Sattigeri, P. and Balakrishnan, A.: Var iational inference of disentangled latent concepts from unlabe led observations, ICLR, 2018.
- [Marimont 2021] Marimont, N. S. and Tarroni, G.: Anomaly dete ction through latent space restoration using vector quantized variational autoencoders, ISBI, 2021.