

Variational AutoEncoder とその拡張手法に基づく異常検知の傾向調査

井上 涼太† 平川 翼† 山下 隆義† 藤吉 弘亘†

† 中部大学

E-mail: inoue@mprg.cs.chubu.ac.jp

1 はじめに

異常検知とは、大多数のデータとは異なる状態のデータを検出する手法である。異物検知や製品の外観検査などの異常検知を自動化するためには、膨大な量の正常データと異常データを用いて教師あり学習を行う。異常データを十分に用意できない場合は、正常データのみから Variational AutoEncoder (VAE) [1] を学習して異常検知する手法が有効である [2][3][4]。VAE は変分推論を軸とした生成モデルであり、中間層で得られる特徴量からデータを生成する。学習データ分布を確率分布で表現できるため、データの補間によって学習データに含まれないデータの生成が可能である。正常データのみを用いて学習を行い、正常データ分布を上手く捉えた VAE は、正常な画像を再構成する。従って、異常データを入力した場合でも正常データとして VAE は再構成するため、入力データと再構成データの類似度を求めることで異常の検知が可能となる。

VAE による異常検知の性能は、VAE の画像生成能力に依存するという課題がある。VAE には大きく 2 つの問題点がある。1 つ目は潜在空間の表現力が弱く、生成画像の多様性がないことである。これは、再構成誤差を最小化するように学習するため、潜在空間が疎になり情報量が少なくなることが原因である。2 つ目は VAE の潜在空間が連続的なベクトルで表現されており、このようなベクトルでは入力画像の特徴を表現しきれずに生成画像の品質が低下することである。これらの問題点を解決するため、beta-VAE [5]、Vector-Quantized (VQ)-VAE [6] 及び Stochastic-Quantization (SQ)-VAE [7] が提案されている。beta-VAE は、再構成誤差に対する正則化の強さを調整するハイパーパラメータ β を導入した手法であり、潜在空間の表現力を高めることで生成画像の多様性を向上させている。VQ-VAE は、連続的な潜在空間を離散的な潜在空間へ量子化した手法であり、生成画像の品質を向上させている。SQ-VAE は、VQ-VAE の量子化方法を改善した手法であり、確率的量子化から徐々に決定論的量子化へ変化させることで VQ-VAE のコードブック崩壊 [8] を抑制している。

VAE を拡張した手法が提案されている一方で、これらのモデルが入力画像と再構成画像の類似度による異常

検知に用いた場合、異常検知性能がどの程度向上するかを統一的な評価が十分に行われていない。そこで本研究では、VAE、beta-VAE、VQ-VAE 及び SQ-VAE を異常検知へ適用し、MVTec Anomaly Detection (MVTec-AD) [9] データセットを用いて、各手法の異常検知における性能を評価する。定量的評価では各モデルの再構成誤差やこれに基づく異常検知性能を、定性的評価では各モデルの再構成の可視化画像を比較して傾向を調査する。

2 関連研究

本章では、VAE とその拡張手法である beta-VAE、VQ-VAE 及び SQ-VAE について述べる。

2.1 VAE

VAE [1] は、入力データ \mathbf{x} の確率分布 $p(\mathbf{x})$ を推定することを目的とした深層生成モデルの一つである。しかし、 $p(\mathbf{x})$ を直接求めることは不可能であるため、データを潜在空間へ埋め込む $q(\mathbf{z}|\mathbf{x};\phi)$ を導入し、 $p(\mathbf{x})$ を任意の潜在空間で近似的に求め、 \mathbf{x} の分布を推定する。ここで、 ϕ は学習パラメータ、 \mathbf{z} は潜在変数である。VAE はエンコーダとデコーダの 2 つのニューラルネットワークから構成される。エンコーダ $q(\mathbf{z}|\mathbf{x};\phi)$ は入力 \mathbf{x} から潜在変数 \mathbf{z} を推定し、デコーダ $p(\mathbf{x}|\mathbf{z};\theta)$ は潜在変数 \mathbf{z} から入力 \mathbf{x} を再構成する。ここで、 θ はデコーダの学習パラメータである。VAE では、エンコーダとデコーダを含むモデルにデータを与えたものについて $p_{\theta,\phi}(\mathbf{x})$ と定義し、 $p_{\theta,\phi}(\mathbf{x})$ を最尤推定により入力 \mathbf{x} を表現する適切なパラメータを探索する。しかし、 $p_{\theta,\phi}(\mathbf{x})$ は連続値を扱うため積分の最大化が必要であり、計算が困難である。そのため VAE では、Evidence Lower Bound (ELBO) を導入する。ELBO を式 (1) に示す。

$$\begin{aligned}\log p(\mathbf{x}) &= \log \int p(\mathbf{x}, \mathbf{z}) d\mathbf{z} \\ &= \log \int q(\mathbf{z}|\mathbf{x}) \frac{p(\mathbf{x}, \mathbf{z})}{q(\mathbf{z}|\mathbf{x})} d\mathbf{z} \\ &\geq \int q(\mathbf{z}|\mathbf{x}) \log \frac{p(\mathbf{x}, \mathbf{z})}{q(\mathbf{z}|\mathbf{x})} d\mathbf{z} \\ &= \mathcal{L}(\mathbf{x}, \mathbf{z})\end{aligned}\tag{1}$$

式 (1) には不等号があるため、下限を最適化した場合求めたい値と ELBO にギャップが生じる。そのため、VAE では $\log p(\mathbf{x}) - \mathcal{L}(\mathbf{x}, \mathbf{z})$ を計算してギャップを求める。Kullback–Leibler (KL) ダイバージェンスを用いてギャップを計算するため、ELBO は式 (2) のように表す。

$$\begin{aligned} \mathcal{L}_{vae} &= \log p(\mathbf{x}) - D_{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z}|\mathbf{x})] \\ &= \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\log p(\mathbf{x}|\mathbf{z})] - D_{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})] \end{aligned} \quad (2)$$

ここで、 D_{KL} は KL ダイバージェンスである。従って VAE の学習は、ELBO の最大化を損失関数とし、エンコーダとデコーダを同時に最適化する。

2.2 beta-VAE

beta-VAE [5] は、入力データの特徴を表現する潜在変数の分布を多変量正規分布に近づけた手法である。VAE の損失関数の正則化項に重み係数 $\beta (> 1)$ を乗算しており、 β を制御することで各潜在変数がより強く事前分布に従うように制約をかけることが可能である。beta-VAE の損失関数を式 (3) に示す。

$$\begin{aligned} \mathcal{L}_{beta} &= \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\log p(\mathbf{x}|\mathbf{z})] \\ &\quad - \beta D_{KL}[q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})] \end{aligned} \quad (3)$$

β の値を大きくすると正則化が強まり潜在変数は分離されるが、再構成誤差は軽視されるため画像がぼやけやすくなる。再構成画像の品質と潜在変数の分離度はトレードオフであるため、 β は慎重に設定しなければならない。しかし、適切な値を設定することでデータの特徴を上手く捉えた潜在変数を得ることができるため、入力データの再現性が向上する。

2.3 VQ-VAE

VQ-VAE [6] は、潜在空間を離散的なベクトルで表現することで表現力を向上させた手法である。通常の VAE では、潜在変数が連続的なベクトルで表現されており、強力なデコーダを使用した場合に潜在変数を見捨てた生成を行う事後分布崩壊が発生する問題がある。VQ-VAE では、潜在変数をコードブックと呼ばれる離散的なベクトルで量子化することでこの問題を回避している。具体的には、エンコーダが出力した潜在変数 \mathbf{z}_e とし、コードブック $\mathbf{B} := \{\mathbf{b}_k\}_{k=1}^K$ の各ベクトルとの間で最も近いベクトル \mathbf{b}_k を求める。ベクトル量子化を式 (4) に示す。

$$\mathbf{z}_q = \mathbf{b}_k, \text{ where } k = \operatorname{argmin}_k \|\mathbf{z}_e - \mathbf{b}_k\|_2 \quad (4)$$

ここで、 \mathbf{z}_q は量子化した潜在変数である。これにより、VQ-VAE は離散的な潜在空間を獲得する。VQ-VAE の損失関数は VAE と同様に ELBO を最大化することである。ただし、量子化の誤差を計算するためにベクトル量子化に関する損失を加える。VQ-VAE の損失関数

を式 (5) に示す。

$$\begin{aligned} \mathcal{L}_{vq} &= \log p(\mathbf{x} | \mathbf{z}_q) + \|\operatorname{sg}[\mathbf{z}_e] - \mathbf{b}\|_2^2 \\ &\quad + \beta \|\mathbf{z}_e - \operatorname{sg}[\mathbf{b}]\|_2^2 \end{aligned} \quad (5)$$

ここで、 $\operatorname{sg}[\cdot]$ は勾配停止演算子であり、 β は [0.1, 2.0] の間で設定される。また、Exponential Moving Average (EMA) 更新を第 2 項に対して適用し、性能と収束率を向上させている。

2.4 SQ-VAE

SQ-VAE [7] は、VQ-VAE の量子化方法を改善した手法であり、コードブック崩壊 [8] を抑制している。コードブック崩壊とは、エンコーダが出力した潜在変数を最近傍のコードブックで量子化するため、使用されるコードブックに偏りが生じる現象である。また VQ-VAE では、勾配停止演算子や EMA 更新などのヒューリスティックな設計を採用している。SQ-VAE では、セルフアニーリングと学習可能な確率的量子化により、ヒューリスティックな設計が不要でありコードブック崩壊を抑制している。

SQ-VAE の量子化は、学習初期では確率的に行われる。そして、学習が進むにつれて決定的量子化へと徐々に変化する。これをセルフアニーリングと呼び、コードブックの利用率が向上する。量子化した潜在変数 \mathbf{z}_q を逆量子化するとき、要素 $z_{q,i}$ を平均、パラメータ Σ_φ を分散としてガウス分布に従うと仮定する。これにより、確率的量子化を導出する。SQ-VAE のデコーダ $p(\mathbf{x}|\mathbf{z}_q)$ の出力にはガウス分布 $\mathcal{N}(f_\theta(\mathbf{z}_q), \sigma^2 I)$ が採用されている。ここで、 $f_\theta(\cdot)$ はネットワークの出力、 $\sigma^2 I$ は学習可能なパラメータである。SQ-VAE の損失関数を式 (6) に示す。

$$\begin{aligned} \mathcal{L}_{sq} &= \mathbb{E}_{q_\omega(\mathbf{z}|\mathbf{x})} \hat{P}_\varphi(\mathbf{z}_q | \mathbf{z}) \left[\frac{1}{2\sigma^2} \|\mathbf{x} - f_\theta(\mathbf{z})\|_2^2 \right. \\ &\quad \left. + \mathcal{R}_\varphi^{\mathcal{N}}(\mathbf{z}, \mathbf{z}_q) \right] \\ &\quad - \mathbb{E}_{q_\omega(\mathbf{z}|\mathbf{x})} H(\hat{P}_\varphi(\mathbf{z}_q | \mathbf{z})) \\ &\quad + \frac{D}{2} \log \sigma^2 + \text{const.} \end{aligned} \quad (6)$$

ここで、 $q_\omega(\mathbf{z} | \mathbf{x})$ はエンコード過程、 $\hat{P}_\varphi(\mathbf{z}_q | \mathbf{z})$ は量子化過程、 $\mathcal{R}_\varphi^{\mathcal{N}}(\mathbf{z}, \mathbf{z}_q)$ は Σ_φ に応じた正則化項である。

3 評価実験

本章では、第 2 章で述べたモデルを入出力類似度による異常検知に適用した際の再構成精度、異常検知精度及び再構成画像を比較する。

3.1 実験条件

本実験では、MVTec-AD [9] を学習・評価用データセットに使用する。MVTec-AD とは、テクスチャを 5 カテゴリとオブジェクトを 10 カテゴリの計 15 カテゴリ

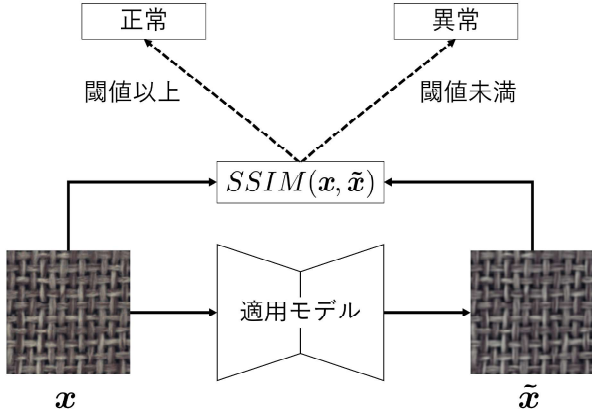


図 1: SSIM による異常検知の流れ

りからなる教師なし学習による異常検知用データセットである。学習用として 3,629 枚、評価用として 1,725 枚が用意されている。本実験では、大規模なデータセットで学習するためデータ増幅により、1 万枚に画像を増幅する。データ増幅は、 512×512 ピクセルにリサイズして、画像内を 128×128 ピクセルでランダムクロップし、左右上下に反転する。

すべての学習において、バッチサイズは 128、学習回数は 200 エポック、学習率は 1.0×10^{-4} 、潜在変数は 64 次元とする。また、beta-VAE の β は 4.0、VQ-VAE と SQ-VAE のコードブックサイズ K は 512 とする。

3.2 SSIM による異常検知

本実験では、入力画像 x と再構成画像 \tilde{x} の類似度を用いて異常であるかを検知する。類似度の計算には Structural Similarity Index Measure (SSIM) [10] を用いる。SSIM による異常検知の流れを図 1 に示す。類似度が閾値以上の場合には正常、閾値未満の場合には異常とする。SSIM とは、画像構造の類似度を $[0, 1]$ で表す指標である。画像構造の類似度とは、画像間の輝度、コントラスト及び構造の 3 要素の乗算から成る。SSIM を式 (7) に示す。

$$SSIM(x_1, x_2) = [l(x_1, x_2)]^\alpha \cdot [c(x_1, x_2)]^\beta \cdot [s(x_1, x_2)]^\gamma$$

$$l(x_1, x_2) = \frac{2\mu_{x_1}\mu_{x_2} + C_1}{\mu_{x_1}^2 + \mu_{x_2}^2 + C_1}$$

$$c(x_1, x_2) = \frac{2\sigma_{x_1}\sigma_{x_2} + C_2}{\sigma_{x_1}^2 + \sigma_{x_2}^2 + C_2}$$

$$s(x_1, x_2) = \frac{\sigma_{x_1 x_2} + C_3}{\sigma_{x_1}\sigma_{x_2} + C_3}$$
(7)

ここで、 x_1, x_2 は比較画像の各画素を要素とするベクトル、 $l(x_1, x_2)$ は輝度、 $c(x_1, x_2)$ はコントラスト、 $s(x_1, x_2)$ は構造、 α, β, γ, C を正の定数、 μ は平均画素、 σ は画素の標準偏差である。

3.3 評価指標

本実験において、再構成精度は SSIM、異常検知精度は Area Under the Receiver Operating Characteristic (AUROC) を用いた評価を行う。

AUROC は、2 値分類で用いられる評価指標である。閾値を変化させた際の True Positive Rate (TPR) と False Positive Rate (FPR) から Receiver Operating Characteristic (ROC) [11] 曲線が算出できる。また、ROC 曲線の下側の面積を Area Under the Curve (AUC) [12] と呼び、1 に近いほど精度が高い。AUROC の縦軸は TPR、横軸は FPR であり、以下の式 (8) から算出する。

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (8)$$

ここで、 TP は True Positive、 FP は False Positive、 TN は True Negative、 FN は False Negative である。

3.4 実験結果

本節では、第 2 章で述べたモデルを SSIM による異常検知に適用した際の再構成精度、異常検知精度及び再構成画像を比較して傾向を調査する。

3.4.1 再構成と異常検知の定量的評価

SSIM と AUROC により算出した再構成精度と異常検知精度を表 1 に、各手法の ROC 曲線を図 2 に示す。ここで、表の値は Seed 値を合計 5 回変更したときの平均、太字は最も高いスコアを示す。また、再構成精度は、正常画像のみを用いて計算する。表 1 より、VAE, beta-VAE よりも VQ-VAE, SQ-VAE の方が再構成精度が高くなる傾向が確認できる。これは、潜在空間を離散的なベクトルで表現することで、潜在変数が持つ特徴の表現力が向上しているためである。beta-VAE が最も低い精度であったのは、設定した値 β が適切ではなく、データの特徴を上手く捉えられなかったと考えられる。表 1 と図 2 より、異常検知精度はテクスチャでは VQ-VAE, SQ-VAE の方が高精度となる傾向があるが、carpet カテゴリと grid カテゴリでは VAE の方が高精度であることが確認できる。また、オブジェクトでは、VAE, beta-VAE の方が高精度となる傾向が確認できる。本実験では入力画像と再構成画像の類似度に基づいて異常検知しているが、再構成精度と異常検知精度に相関がない。これは、VQ-VAE, SQ-VAE の画像生成能力が強力なため、異常画像を正常画像に変換できずに誤検知を誘発したと考えられる。

3.4.2 再構成の定性的評価

carpet カテゴリ、tile カテゴリ、bottle カテゴリ及び cable カテゴリの再構成画像を図 3 に示す。図 3 より、

表 1: 再構成精度と異常検知精度の比較

category	再構成精度 ↑				異常検知精度 ↑			
	VAE	beta-VAE	VQ-VAE	SQ-VAE	VAE	beta-VAE	VQ-VAE	SQ-VAE
carpet	0.585	0.303	0.965	0.936	0.836	0.782	0.769	0.808
grid	0.408	0.369	0.853	0.748	0.746	0.742	0.608	0.645
leather	0.001	0.000	0.586	0.746	0.465	0.494	0.699	0.708
tile	0.245	0.153	0.956	0.929	0.547	0.532	0.834	0.810
wood	0.143	0.007	0.634	0.684	0.560	0.495	0.747	0.702
テクスチャの平均	0.276	0.137	0.798	0.808	0.630	0.576	0.731	0.734
bottle	0.588	0.600	0.748	0.746	0.859	0.863	0.678	0.700
cable	0.515	0.472	0.921	0.867	0.736	0.729	0.671	0.643
capsule	0.749	0.746	0.765	0.462	0.729	0.730	0.548	0.350
hazelnut	0.356	0.355	0.426	0.319	0.796	0.803	0.353	0.274
metalnut	0.451	0.444	0.641	0.574	0.596	0.611	0.318	0.344
pill	0.360	0.349	0.531	0.429	0.482	0.487	0.392	0.373
screw	0.394	0.381	0.468	0.198	0.572	0.585	0.323	0.216
toothbrush	0.404	0.388	0.434	0.341	0.409	0.399	0.351	0.342
transistor	0.822	0.801	0.921	0.813	0.851	0.851	0.642	0.601
zipper	0.179	0.168	0.455	0.081	0.663	0.679	0.612	0.458
オブジェクトの平均	0.481	0.470	0.631	0.483	0.669	0.673	0.488	0.430

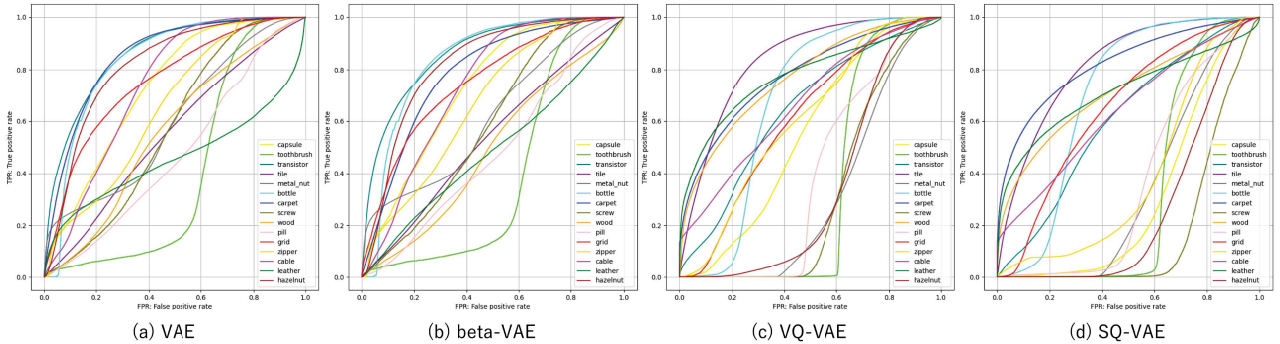


図 2: 各手法の ROC 曲線

VAE, beta-VAE の再構成画像は低品質ではあるが、異常画像を正常画像として再構成している。一方で、VQ-VAE, SQ-VAE の再構成画像は高品質ではあるが、異常画像をそのまま再構成していることが確認できる。学習用データセットには異常画像は含まれていないため、各モデルが獲得している潜在空間は正常画像を表現している。連続的な潜在空間を持つ VAE と beta-VAE は、入力画像を連続的なベクトルで写像するため、異常を正常に復元することができる。しかし、離散的な潜在空間を持つ VQ-VAE と SQ-VAE は、入力画像を離散的なベクトルであるコードブックで表現する。そのため、コードブックの組み合わせ次第では学習していない特徴も表現できると考えられる。以上のことから、入力画像と再構成画像類似による異常検知では、VQ-VAE や SQ-VAE のように離散的な潜在空間を持つモデルではなく、VAE や beta-VAE のように異常を正常に復元することができる連続的な潜在空間を持つモデルが有効である。

間を持つモデルではなく、VAE や beta-VAE のように異常を正常に復元することができる連続的な潜在空間を持つモデルが有効である。

3.4.3 コードブックの可視化

離散的な潜在空間を持つモデルは、入力画像と再構成画像類似度による異常検知に適用することが困難である。そこで、コードブックの種類や選択時の距離を可視化して、正常と異常を分離できるか確認する。VQ-VAE に carpet カテゴリと tile カテゴリを入力したときのコードブックの種類と距離をヒートマップで可視化した結果を図 4 に、正常、色異常及び形状異常の距離平均を表 2 に示す。ここで、図 3 の異常とは入力画像の異常領域を白で示す。また、図 3 の種類とは 512 種類のコードブックを示しており、同じコードブックには

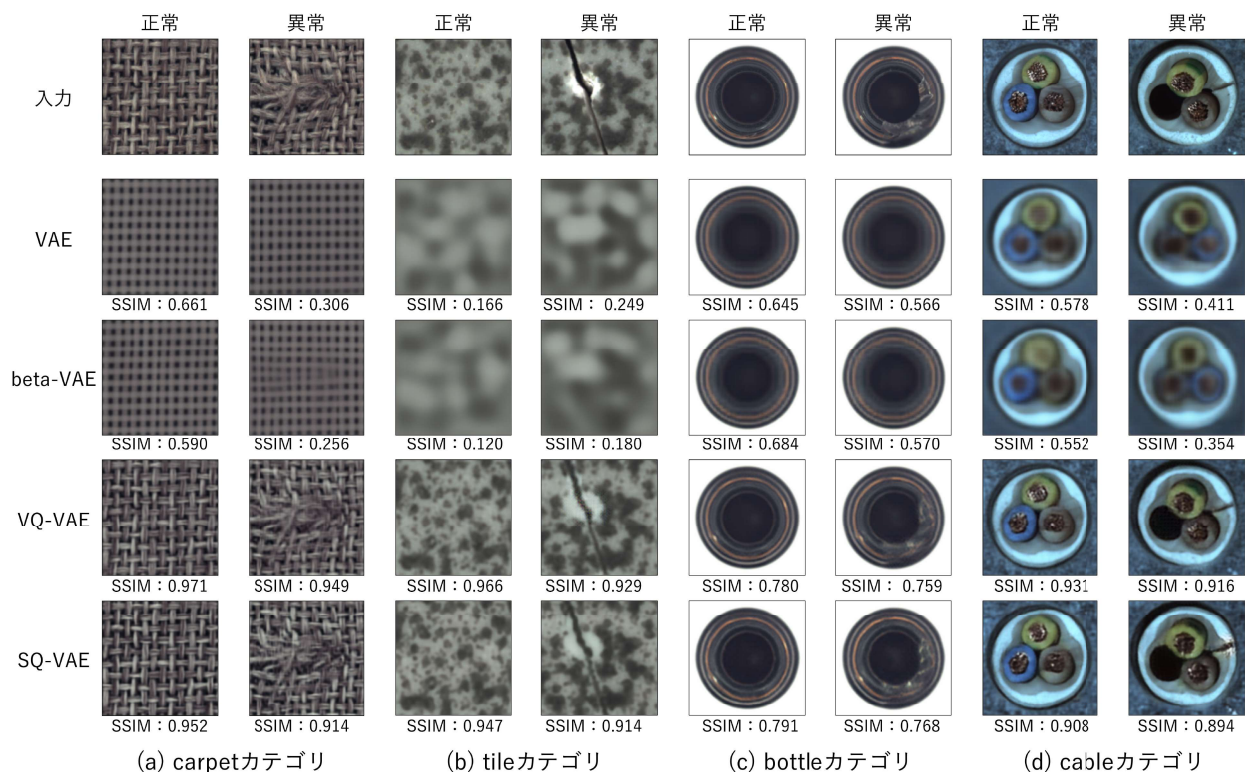


図 3: 各手法の再構成画像の比較

表 2: コードブックの距離平均

	正常	色異常	形状異常
carpet	0.0020	0.0023	0.0023
tile	0.0001	0.0075	0.0900

同じ色を割り当てている。図 3 の距離とは潜在変数を量子化するとき用いた距離であり、距離が小さい場合は青色、大きい場合は赤色で示す。図 4 より、割り当てられるコードブックが正常と異常で異なる傾向が確認できる。特に、正常画像には存在しない色の異常には、正常と異なるコードブックが割り当てられている。しかし、正常と類似した色の異常では割り当てられるコードブックに違いがあまり確認できない。図 4 と表 2 より、carpet カテゴリの異常は正常と類似した色をしているため、正常の距離平均と類似している。しかし、tile カテゴリのように正常に存在しない色の異常の場合は、異常の距離平均が正常とは大きく異なることが確認できる。SQ-VAE においても同様の傾向を確認した。以上のことから、離散的な潜在空間を持つモデルを異常検知に適用する場合、色の異常検知にのみ適用し形状の異常検知は別の手法やモデルを適用すると、モデル単体で適用するよりも精度が向上すると考えられる。

4 おわりに

本研究では、VAE, beta-VAE, VQ-VAE 及び SQ-VAE を入力画像と再構成画像類似度による異常検知に用いた場合において、異常検知精度がどの程度向上するかを統一的な評価で比較した。定量的評価において、再構成精度は SSIM, 異常検知精度は AUROC を用いて評価した。定性的評価において、再構成とコードブックを可視化した。

定量的評価の結果、再構成精度は離散的な潜在空間を持つ VQ-VAE, SQ-VAE が高精度となる傾向を確認した。異常検知度はテキストとオブジェクトで傾向が異なっており、テキストでは VQ-VAE, SQ-VAE が高精度、オブジェクトでは VAE, beta-VAE が高精度となる傾向を確認した。これは、オブジェクトの方が繰り返しパターンが単純なため、連続的な潜在空間を持つモデルの方が適していることが考えられる。従って、テキストのような不規則なパターンのデータには、離散的な潜在空間を持つモデルが適していることが考えられる。

定性的評価の結果、VAE, beta-VAE の再構成画像は低品質ではあるが、異常画像を正常画像として再構成している一方で、VQ-VAE, SQ-VAE の再構成画像は高品質ではあるが、異常画像をそのまま再構成していることを確認した。VQ-VAE のコードブック可視化では、正常に存在しない色の異常に割り当てられるコー

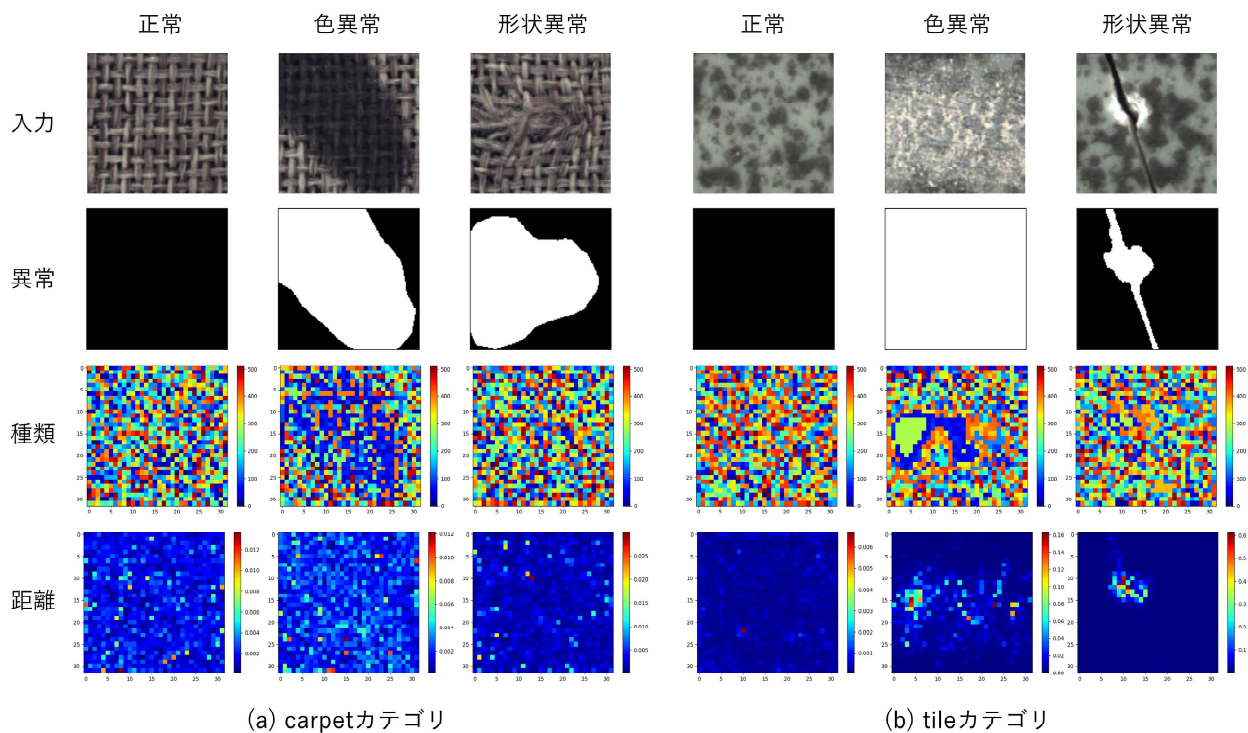


図 4: VQ-VAE のコードブック可視化

ドブックや距離が大きく異なることを確認した。

以上のことから、入力画像と再構成画像類似度による異常検知においてはデータの性質に応じたモデルを選択することが重要であると考えられる。今後は、他の生成モデルによる異常検知における比較調査を行う。

参考文献

- [1] D. P. Kingma and M. Welling: “Auto-Encoding Variational Bayes”, International Conference on Learning Representation, pp. 1–14 (2014).
- [2] D. Dehaene, O. Frigo, S. Combexelle and P. Eline: “Iterative energy-based projection on a normal data manifold for anomaly localization”, International Conference on Learning Representations, pp. 1–17 (2020).
- [3] E. Prifti, J. P. Buban, A. S. Thind and R. F. Klie: “Variational convolutional autoencoders for anomaly detection in scanning transmission electron microscopy”, *Small*, **19**, 16, p. 2205977 (2023).
- [4] D. Kumar, C. Verma, Z. Illes, A. Mittal, B. Bakariya and S. Goyal: “Anomaly detection in chest x-ray images using variational autoencoder”, 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Vol. 6IEEE, pp. 216–221 (2023).
- [5] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed and A. Lerchner: “beta-vae: Learning basic visual concepts with a constrained variational framework”, International Conference on Learning Representations (2016).
- [6] A. van den Oord, O. Vinyals and k. kavukcuoglu: “Neural discrete representation learning”, Advances in Neural Information Processing Systems (Eds. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett), Vol. 30, Curran Associates, Inc. (2017).
- [7] Y. Takida, T. Shibuya, W. Liao, C.-H. Lai, J. Ohmura, T. Uesaka, N. Murata, S. Takahashi, T. Kumakura and Y. Mitsufuji: “SQ-VAE: Variational Bayes on discrete representation with self-annealed stochastic quantization”, Proceedings of the 39th International Conference on Machine Learning (Eds. by K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu and S. Sabato), Vol. 162 of Proceedings of Machine Learning Research, PMLR, pp. 20987–21012 (2022).
- [8] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford and I. Sutskever: “Jukebox: A generative model for music”, arXiv preprint arXiv:2005.00341 (2020).

- [9] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger and C. Steger: “The mvtec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection”, *International Journal of Computer Vision*, **129**, 4, pp. 1038–1059 (2021).
- [10] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli: “Image quality assessment: from error visibility to structural similarity”, *IEEE Transactions on Image Processing*, **13**, 4, pp. 600–612 (2004).
- [11] J. N. Mandrekar: “Receiver operating characteristic curve in diagnostic test assessment”, *Journal of Thoracic Oncology*, **5**, 9, pp. 1315–1316 (2010).
- [12] J. Myerson, L. Green and M. Warusawitharana: “Area under the curve as a measure of discounting”, *Journal of the Experimental Analysis of Behavior*, **76**, 2, pp. 235–243 (2001).