

1. はじめに

工場の生産ラインでは、製品が良品か不良品かを判定する検査を行う。検査を自動化する際、不良品（異常）が少ない場面では、正解ラベルを必要しない教師なし学習による異常検知が有効である。このような異常検知手法として、生成モデルベースの Variational Autoencoder (VAE) が用いられることがある。しかしながら、VAE による画像の再構築は、異常を含む画像が入力されたとき、異常部分を完全に除去できないという問題点がある。そこで、学習済みの VAE にエネルギー関数の勾配をかけ異常部分のみを抽出し、正常画像を再構築するまで更新する手法 (VAE-grad) [1] が提案されている。VAE-grad は学習データの再構成誤差の最小値を基に更新を終了する。VAE の学習が上手くできていない場合、更新が終了しないため、正常画像を変化させてしまう問題がある。そこで本研究では、画像の更新回数に着目し、更新画像や乗じられる勾配の大きさから入力が正常画像の場合は、少ない更新回数で終了し、異常画像の場合は正常に再構築されると更新を終了する方法を提案する。

2. 勾配を利用した再構築

VAE-grad を用いた再構築の流れを図 2 に示す。この手法は、学習済みの AE に基づくエネルギー関数とその勾配を用いて入力画像を再構築する。エネルギー関数 E を式 (1) に示す。ここで、 x_0 は入力画像、 x_t は生成画像、 L_r は VAE の損失関数を表す。また、第 2 項は生成画像が初期画像から離れすぎないようにする制約である。

$$E(x_t) = L_r(x_t) + \lambda \cdot \|x_t - x_0\| \quad (1)$$

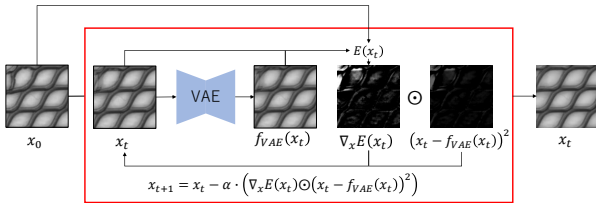


図 1：勾配を利用した入力画像の更新 (VAE-grad)

勾配には、VAE に入力する生成画像 x_t と再構築した画像 $f_{VAE}(x_t)$ の差分を乗算する。これにより、正常部分の更新を抑制し、異常部分の更新を促す。最終的な入力画像の更新方法を式 (2) に示す。この更新を繰り返し、再構成誤差を用いた閾値処理によって更新を終了する。

$$x_{t+1} = x_t - \alpha \cdot (\nabla_x E(x_t) \odot (x_t - f_{VAE}(x_t))^2) \quad (2)$$

3. 提案手法

本研究では、VAE-grad において画像の更新を終了する閾値を決定する 2 つの手法を提案する。1 つ目は、更新画像の類似度に着目し更新を終了する。処理内容を以下に示す。

- ① 生成画像 x_t と再構築画像 $f_{VAE}(x_t)$ の類似度を比較
- ② 類似度が最大のとき更新を終了

2 つ目は、エネルギー関数の勾配に着目する。図 2 は正常画像と異常画像の勾配を示す。正常画像は広範囲に小さな勾配が乗じられているのに対し、異常画像は局所的に大きな勾配が乗じられる。この特徴を活かし更新を終了する。処理方法を以下に示す。

- ① 勾配の最大値 $\max(\nabla_x E(x_t))$ が閾値を下回ったとき更新終了

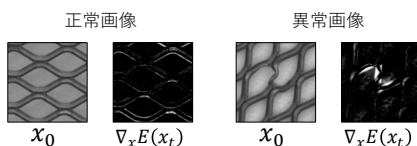


図 2：正常画像と異常画像の勾配

4. 評価実験

画像の更新を終了させる 2 つの閾値処理の有効性を示すために、AUROC 比較および、入力が正常画像と異常画像での更新回数の変化を調査し、最適な閾値処理を模索する。

4.1. 実験条件

データセットに MVTec-AD、評価指標に AUROC を用いる。MVTec-AD は、Texture の 5 カテゴリ、Object の 10 カテゴリで構成される。AUROC は、異常と正常を正確に予測できた割合を表す。更新画像の類似度による閾値処理は、SSIM を導入し VAE に入力する生成画像 x_t と再構築画像 $f_{VAE}(x_t)$ の類似度を算出する。勾配による閾値処理は、勾配の最大値が 0.2 以下で更新を終了させる。エポック数は 300、潜在変数の次元数は 100、画像の最大更新回数は 300 回、学習率は 1×10^{-4} 、 α は 0.5、 λ は 0.05 とする。

4.2. 実験結果

各閾値処理を用いた場合の AUROC を表 1 に示す。表 1 では、Texture と Object の 1 部のカテゴリを示す。Texture では AUROC の差は小さいが、Object では勾配による閾値処理の方が AUROC が高く、平均を見ると 0.9 を越えており、閾値処理が上手く作用している。

表 1：AUROC の比較

category	閾値：類似度	閾値：勾配
grid	0.928	0.948
tile	0.568	0.554
wood	0.844	0.859
Texture 平均	0.831	0.835
metanlut	0.929	0.940
toothbrush	0.929	0.949
Object 平均	0.882	0.902

4.3. 2 つの閾値処理の更新回数

各閾値処理を用いた正常画像と異常画像の更新回数を表 2 に示す。表 2 は、表 1 と同様のカテゴリでの閾値処理別の更新回数を示している。類似度による閾値処理は、正常画像も異常画像も更新終了にかかる更新回数が多い。勾配による閾値処理は、正常画像は異常画像よりも少ない更新回数で更新が終了している。

表 2：更新回数の比較

category	閾値：類似度		閾値：勾配	
	正常	異常	正常	異常
grid	297.5	299.9	130.9	164.0
tile	300	296.4	282.2	242.8
wood	278.1	280.5	68.3	125.4
Texture 平均	193.1	295.1	127.4	146.8
matalnut	299.7	299.8	110.5	151.1
toothbrush	300	298.1	98.1	107.9
Object 平均	297.2	297.8	169.1	192.9

5. おわりに

本研究では、VAE-grad の画像の更新を終了させる閾値処理を導入した。実験結果より、勾配の大きさにより更新を終了させる手法では、正常画像は異常画像よりも少ない更新回数で更新を終了しており、AUROC も向上することを確認した。今後は生成画像に着目し、再構築が困難であるランダム性のある模様を上手く再構築できるような VAE の学習を行い精度向上を目指す。

参考文献

- [1] D. Dehaene, *et al.*, “Iterative energy-based projection on a normal data manifold for anomaly localization”, ICLR, 2020.