Center Loss を導入した VAE による埋め込み空間における異常検知

EP19016 井上涼太　　指導教授：山下隆義

1. はじめに

正常データのみを用いて学習した Variational Autoencoder (VAE) [1] は，入力に対応した正常データの再構成が可能である．異常データを入力した場合は，正常データに近い画像が再構成され，入出力の類似度が小さいときに異常と検知する．このとき，VAE が再構成に失敗すると，誤検知を誘発する．そこで本研究では，入出力の類似度だけでなく，埋め込み空間における異常検知手法を提案する．埋め込み空間では，学習データから求めた重心を基準として異常データを検知する．また，異常検知の対象となる埋め込み空間のバラつきを Center Loss [2] によって抑制することで，更なる異常検知精度の向上を図る．

2. VAE を用いた異常検知

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が生成できる．正常データの特徴を上手く捉えた VAE は，正常な特徴のみ再構成するため，異常データを入力した場合でも正常データとして再構成される．従って，入力データの類似度を求めることで異常データを特定できる．

3. 提案手法

埋め込み空間における異常検知を入出力の類似度による異常検知と組み合わせることで，異常検知精度の向上を図る．提案手法による異常検知の流れを図 1 に示す．

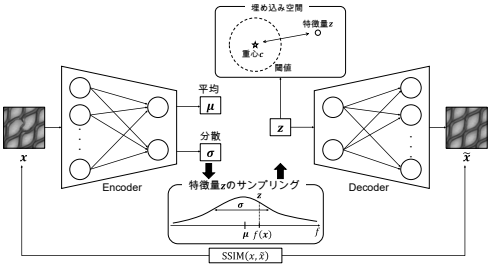

図 1：提案手法による異常検知

3.1. 埋め込み空間における異常検知

埋め込み空間における異常検知は，重心 $\boldsymbol{c}$ とテストデータの特徴量 $\boldsymbol{z}$ とのユークリッド距離 $d(\boldsymbol{z}, \boldsymbol{c})$ を求め，閾値を超える場合は異常とする．重心 $\boldsymbol{c}$ は，VAE に入力して得た全ての学習データの特徴量を平均して求める．

3.2. 画像空間と埋め込み空間を組み合わせた異常検知

従来の画像空間と 3.1. の埋め込み空間を用いて判定する．その流れを以下に示す．

Step1 VAE へ画像 $\boldsymbol{x}$ を入力して画像空間と埋め込み空間，それぞれで異常検知を行う．

$$\mathrm{pred}_S = \mathbf{1}[\mathrm{SSIM}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) > T_S]$$
$$\mathrm{pred}_D = \mathbf{1}[d(\boldsymbol{z}, \boldsymbol{c}) > T_D]$$

ここで，$\tilde{\boldsymbol{x}}$ は再構成画像，T_S は画像空間の閾値，T_D は埋め込み空間の閾値，$\mathbf{1}$ はインジケーター関数である．

Step2 採用する結果を決定するため，同じ入力から得られた結果 pred_S，pred_D を比較する．

Step3 結果が一致する場合はその結果を採用する．結果が異なる場合は類似度とユークリッド距離において閾値から離れている結果を採用する．従って，以下のように求める．

$$\mathrm{Output} = \begin{cases} \mathrm{pred}_S \text{ if } \mathrm{SSIM}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) - T_S > d(\boldsymbol{z}, \boldsymbol{c}) - T_D \\ \mathrm{pred}_D \text{ otherwise} \end{cases}$$

3.3. Center Loss を導入した VAE

正常データの重心位置を安定させるため，Center Loss によって埋め込み空間のバラつきを抑制する．VAE の特徴量から Center Loss を計算し，類似した特徴をまとめる．提案手法の損失関数は以下のようになる．

$$\mathcal{L} = \mathbb{E}_{q(\boldsymbol{z}|\boldsymbol{x})}[\log p(\boldsymbol{x}|\boldsymbol{z})] - D_{KL}[q(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z})] + \lambda \frac{1}{2} \sum_{i=1}^{m} \| \boldsymbol{z}_i - \boldsymbol{c}_y \|_2^2$$

ここで，λ はハイパーパラメータ，m はバッチサイズ，y はクラスラベル，$\boldsymbol{c}_y$ はクラス重心である．

4. 評価実験

本実験では，異常検知精度を比較することで，提案手法の有効性を示す．

4.1. 実験条件

本実験では，MVTec-AD データセットを用いて VAE と提案手法を比較する．MVTec-AD は，異常検知用データセットであり，5 種類の Texture と 10 種類の Object クラスから構成される．本実験では，5 種類の Texture クラス，それぞれで学習と異常検知を行う．学習時の設定はバッチサイズを 128，学習回数を 300，学習率を 1.0×10^{-4}，潜在変数を 25 次元，重心数を 1，$\lambda = 1$ とする．

4.2. 異常検知精度

表 1 に従来手法と提案手法の検出率を示す．ここで，Ours1 は画像空間と埋め込み空間を組み合わせた異常検知による結果である．Ours2 は Ours1 に Center Loss を組み込んだ VAE による結果である．表 1 より，埋め込み空間による異常検知を組み合わせることで精度が向上している．これにより，埋め込み空間を活用した提案手法の有効性を示せた．

表 1：検出率の比較

Category	従来手法	Ours1	Ours2
carpet	0.724	0.698	**0.737**
grid	0.636	0.621	**0.695**
leather	0.500	0.527	**0.541**
tile	0.615	**0.698**	0.680
wood	0.601	0.664	**0.664**
平均	0.615	0.641	**0.663**

4.3. 考察

提案手法により正しく異常検知した例を図 2 に示す．ここで，黄色のマスクは異常箇所である．図 2 より，細かい異常箇所が広範囲に分布しており，テクスチャが不規則であることが確認できる．再構成画像全体がぼやけて正しくテクスチャを再構成できていないため，従来手法は誤検知する．一方，埋め込み空間による異常検知では重心との距離が大きくなるため，正常に検知できたと考えられる．Center Loss によって埋め込み空間のバラつきを抑制すると精度が向上することから，Center Loss を VAE に組み込むことは異常検知において有効であると考えられる．

入力画像　　再構成画像

図 2：提案手法により正しく異常検知したテストデータ

5. おわりに

本研究では，埋め込み空間における異常検知を入出力の類似度による異常検知と組み合わせた手法を提案した．提案手法は従来手法よりも異常検知精度が向上することを確認した．また，VAE に Center Loss を導入することで，異常検知精度の更なる向上を確認した．今後は，VAE を学習する際に使用した Center Loss の重心を使用した異常検知を検討する．

参考文献

[1] D. P. Kingma, *et al.*, “Auto-Encoding Variational Bayes”, ICLR, 2014.

[2] Y. Wen, *et al.*, “A Discriminative Feature Learning Approach for Deep Face Recognition”, ECCV, 2016.