

深層生成モデルによる非正則化異常度を用いた工業製品の異常検知

Anomaly Manufacturing Product Detection using Unregularized Anomaly Score on Deep Generative Models

立花亮介*¹ 松原崇*¹ 上原邦昭*¹
Ryosuke Tachibana Takashi Matsubara Kuniaki Uehara

*¹神戸大学 大学院 システム情報学研究科
Graduate School of System Informatics, Kobe University

One of the most common needs in manufacturing plants is rejecting products not coincident with the standards as anomalies. Manufacturing companies usually employ numerous inspectors for anomaly detection and it takes a high cost. Accurate and automatic anomaly detection reduces inspection cost and improves product reliability. In unsupervised anomaly detection, a probabilistic model detects test samples with lower likelihoods as anomalies. Recently, a probabilistic model called deep generative model (DGM) has been proposed for end-to-end modeling of natural images and already achieved a certain success. However, anomaly detection of machine components with complicated structures is still challenging because they produce a wide variety of the normal image patches with lower likelihoods. For overcoming this difficulty, we propose unregularized score for the DGM. As its name implies, the unregularized score is the anomaly score of the DGM without the regularization terms. The unregularized score is robust to the inherent complexity of a sample and has a smaller risk of rejecting a sample appearing less frequently but being coincident with the standards. Experimental results of anomaly detection on the real manufacturing datasets show better performance of the unregularized score compared to existing approaches.

1. はじめに

工業製品の製造現場において最も重要であることの一つは、製造された工業製品が期待される仕様を満たしていないときに、その製品を不良品であるとみなして取り除くことである。現状では異常な製品の発見や除去は人手で行うことが一般的であり、企業は異常製品の除去に対して高い人的コストを支払っている。このような状況において、正確かつ自動的な異常検知によって製品の信頼性を高めると同時に、製品の点検にかかるコストを削減することが求められている。

教師なし学習における異常検知では、確率モデルによってデータの負の対数尤度を求め、その値を異常度としてこれが高いものを異常とみなす。従来では、データから主成分分析 (Principal Components Analysis; PCA) などで特徴抽出を行った後、得られた特徴を用いて混合ガウスモデル (Gaussian Mixture Model; GMM) などの確率モデルを学習する [Saligrama 12]。最近では、画像から直接尤度を求めることが出来る深層生成モデル (Deep Generative Model; DGM) と呼ばれる確率モデルが提案されており、異常検知において一定の成果を達成している [Ribeiro 17]。しかしながら、工業製品はその構造が複雑かつ多様であるため製品を撮影した際に画像内における各部分の出現頻度が異なり、尤度が必ずしも異常度に対応せず、正しく評価できないという問題がある。

このような問題の解決のため、本論文では深層生成モデルにおいて、非正則化異常度を用いた異常検知手法を提案する。非正則化異常度という名前が示すように、これは深層生成モデルにおける異常度、つまり負の対数数度から正則化項を取り除いたものである。非正則化異常度はデータが潜在的に含有する複雑さに対して堅牢であり、画像内における各部分の出現頻度に依存することなく評価を行える。提案手法の有効性を検証するために、工業製品の画像データに対して本手法を適用させ、異常検知性能においてその有効性を既存の手法と比較する。

2. 非正則化異常度

確率的モデルである VAE や GMM における従来の異常度である負の対数尤度は、いくつかの意味をもつ項に分割できる。提案手法である非正則化異常度は、その項の一つを取り出し新たな異常度として再定義したものである。ここでは、VAE 及び GMM の異常検知における従来の異常度から、提案手法である非正則化異常度の導出を行う。

VAE はパラメータ θ で表現される生成モデル $p_\theta(x)$ に対してモデルエビデンスの下限 (ELBO) を最大化することで学習を行うモデルである。VAE は入力 x を潜在変数 z に変換する符号化器及び潜在変数 z から入力の再構成を行う復号化器から構成される。変分事後分布 $q_\phi(z|x)$ 及び条件付き確率 $p_\theta(x|z)$ は、共に分散共分散行列が対角行列となっている多変量正規分布としてモデル化される。符号化器は入力 x を受け取った後、変分事後分布 $q_\phi(z|x)$ における平均ベクトル μ_z と標準偏差ベクトル σ_z の二つの要素の出力を行う。復号化器は、条件付き確率 $p_\theta(x|z)$ における平均ベクトル μ_x と標準偏差ベクトル σ_x の二つの要素の出力を行う。変分事後分布 $q_\phi(z|x)$ のモンテカルロサンプリングを行う代わりに潜在変数 z の最大事後確率推定 μ_z が異常検知において使用される。よって、負の ELBO $\mathcal{L}_{VAE}(x)$ は以下の式 (1) によって表される。

$$\begin{aligned} \mathcal{L}_{VAE}(x) &= D_{KL}(q_\phi(z|x)||p(z)) - \log p_\theta(x|\mu_z) \\ &= D_{VAE}(x) + A_{VAE}(x) + M_{VAE}(x). \end{aligned} \quad (1)$$

ただし、 $D_{VAE}(x)$, $A_{VAE}(x)$, $M_{VAE}(x)$ はそれぞれ

$$\begin{aligned} D_{VAE}(x) &= \sum_{j=1}^{N_z} \frac{1}{2} (-\log \sigma_{z_j}^2 - 1 + \sigma_{z_j}^2 + \mu_{z_j}^2), \\ A_{VAE}(x) &= \sum_{i=1}^{N_x} \frac{1}{2} \log 2\pi\sigma_{x_i}^2 \Big|_{z=\mu_z}, \\ M_{VAE}(x) &= \sum_{i=1}^{N_x} \frac{1}{2} \frac{(\mu_{x_i} - x_i)^2}{\sigma_{x_i}^2} \Big|_{z=\mu_z}, \end{aligned} \quad (2)$$

連絡先: 松原崇, 神戸大学大学院システム情報学研究科,
matsubara@phoenix.kobe-u.ac.jp

である。また、 i と j はそれぞれデータ x と潜在変数 z の要素の番号を示す。ニューラルネットワークの見地からみると、 $A_{VAE}(x) + M_{VAE}(x)$ は再構成誤差を表しており、また $D_{VAE}(x)$ は正則化項を表している。 $A_{VAE}(x)$ は正規化定数の対数を取ったものに相当し、これは正規分布 $p_\theta(x|z)$ の確率密度関数の積分を 1 にするものである。 $M_{VAE}(x)$ は式の形としてはマハラノビス距離、あるいはユークリッド距離を正規化したものに近いことがわかる。

ここで、導出された $\mathcal{L}_{VAE,M}(x) = M_{VAE}(x)$ を、従来の負の ELBO $\mathcal{L}_{VAE}(x) = D_{VAE}(x) + A_{VAE}(x) + M_{VAE}(x)$ の代わりとして、VAE による異常検知における非正則化異常度とする。非正則化異常度という名前は、このように従来の異常度から正規化定数の対数 $A_{VAE}(x)$ 及び正則化項 $D_{VAE}(x)$ を取り除いていることに由来する。

GMM において、データ x が属する隠れクラス z は最大事後確率推定によって決定することができる。このときのデータ x の負の対数尤度は以下の式 (3) で表される。

$$-\log w_k p(x|z=k) = D_{GMM}(x) + A_{GMM}(x) + M_{GMM}(x). \quad (3)$$

ただし、 $D_{GMM}(x)$ 、 $A_{GMM}(x)$ 、 $M_{GMM}(x)$ はそれぞれ

$$\begin{aligned} D_{GMM}(x) &= -\log w_k, \\ A_{GMM}(x) &= \frac{1}{2} \log (2\pi)^{N_x} |\Sigma_k|, \\ M_{GMM}(x) &= \frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k), \end{aligned} \quad (4)$$

である。但し、 $k = \arg \max_z p(z|x)$ である。そして、導出された $\mathcal{L}_{GMM,M} = M_{GMM}(x)$ を、従来の負の対数尤度 $-\log \sum_{k=1}^{N_z} w_k p(x|z=k)$ の代わりとして、GMM による異常検知における非正則化異常度とする。ここでもし GMM における隠れクラスの数 N_z を 1 とすると、非正則化異常度 $\mathcal{L}_{GMM,M} = M_{GMM}(x)$ は明らかにマハラノビス距離 $(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)$ を 2 で割ったものと等しくなる。

3. 非正則化異常度の解釈

GMM における $D_{GMM}(x)$ 、 $A_{GMM}(x)$ 、and $M_{GMM}(x)$ と、VAE における $D_{VAE}(x)$ 、 $A_{VAE}(x)$ 、and $M_{VAE}(x)$ は、それぞれの式の意味において対応していると考えられる。

工業製品は平らな表面、曲がった部分、ネジの穴など、様々な種類の要素から構成されている。このような場合、図 1 に示すように違う種類の要素グループに属する正常なデータ同士の違いの大きさが、同じ種類の要素のグループに属する正常なデータと異常なデータの違いの大きさよりも大きくなってしまふと考えられる。GMM によって、隠れクラスに各要素グループからの正常なデータが属することになる。そして、異常なデータは特定の異常クラスとして分離されてそれぞれ属するのではなく、多くのデータが属しているのと同じ隠れクラスに属することになる。このような仮定のもとでは、混合係数 w_k は対応しているグループ z にデータが属す頻度を表していることになり、また分散共分散行列 Σ_k は対応しているグループ z からのデータの分布を表していることになる。よって、混合係数の負の対数 $D_{GMM}(x)$ と正規化定数の対数 $A_{GMM}(x)$ は、異常を上手く捉えることができないと考えられる。一方で、マハラノビス距離を 2 で割ったものに等しい $M_{GMM}(x)$ については、分散共分散行列 Σ_k によって正規化が行われたグループ z の分布の中心からの距離を表現しているため、グループ z 内部における異常度として機能することが期待できる。

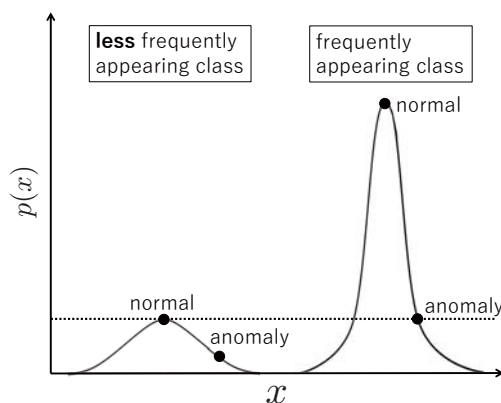


図 1: 工業製品画像における尤度の直感的な図解

VAE を PCA のように教師なし特徴抽出手法として利用すると、潜在変数 z は抽出された特徴として取り扱うことができ、正則化項 $D_{VAE}(x)$ は異常度として機能することが期待できる。しかしながら、VAE は一般的に $N_z < N_x$ であり入力次元よりも出力次元の方が次元数が低い。これは、高次元のデータ空間 \mathcal{X} におけるデータ x を、低次元の潜在空間 \mathcal{Z} へと写像しているということである。入力したデータ x を再構成するためには、VAE は与えられたデータにおいて最も顕著に現れる特徴を学習し、逆に頻度の少ない特徴は無視しなければならない。つまり、異常データ x_a が微小な異常を含んでいたとしても、それが他の部分において同じ種類の要素から取った正常データ x と似通っている場合は、潜在空間 \mathcal{Z} における異常データ x_a の写像 z は正常データ x の写像の近くに分布することになる。このことから、正則化項 $D_{VAE}(x)$ は、混合係数の負の対数 $D_{GMM}(x)$ と同じように異常を検知することに関して有効であるとは言えない。

VAE がデータ x を正確に再構成することが出来る、つまり平均ベクトル μ_x を x に限りなく近づけることが出来るのであれば、標準偏差ベクトル σ_x はほぼ 0 に近い値となるため、これは正規化定数の対数 $A_{VAE}(x)$ を最小化することになる。しかしながら、VAE は最適な平均ベクトル μ_x を推定することが出来ないため、正規化距離 $M_{VAE}(x)$ の値が発散することを防ぐために、標準偏差ベクトル σ_x の値は小さくなくてはならなくなる。VAE では、再構成の不確かさに関して標準偏差 σ_x が調整されることにより、 $A_{VAE}(x)$ と $M_{VAE}(x)$ の二つの基準のバランスが取れるように学習が行われる。このことは、複雑な要素から取られた正常データ x に対して、正規化定数の対数 $A_{VAE}(x)$ が大きな値をとるということの意味する。言い換えれば、標準偏差 σ_{x_i} が潜在的にデータ x の不確かさや本質的な複雑さを表現していることになり、またデータ x が属するグループの不確かさや本質的な複雑さを表現していることにもなる。これは、GMM における分散共分散行列 Σ_k と類似する点である。つまり、正規化定数の対数 $A_{VAE}(x)$ は異常検知では有効であるとは言えない。

ここでの結論としては、正規化距離である $M_{VAE}(x)$ のみが直接的に再構成誤差と関係しているということである。この距離 $M_{VAE}(x)$ は標準偏差 σ_{x_i} によって正規化されている。つまり、正規化距離 $M_{VAE}(x)$ はデータ x の不確かさや本質的な複雑さに対して堅牢であり、これは異常検知を行うにあたって有用な値である。以降の章では、VAE による異常検知において、正規化距離 $M_{VAE}(x)$ を非正則化異常度として実験を行い、異常検知における手法の有効性を検証する。

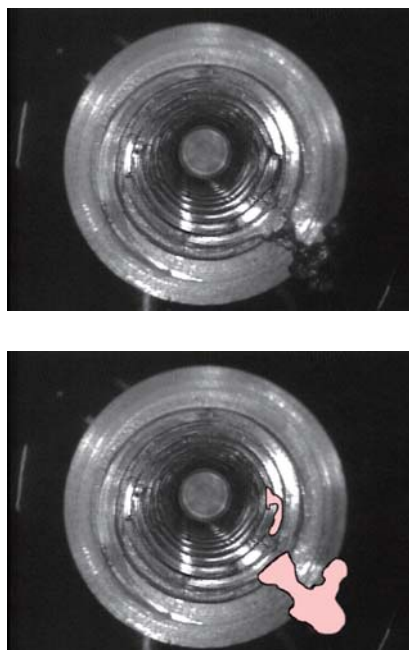


図 2: 異常データと異常箇所の ground truth.

4. 評価実験と考察

4.1 データセット

ネジ穴データセット (The screw dataset) は、ネジの穴を撮影したモノクロ画像のデータセットであり、画像サイズは 640×480 である。各画像の中央部にはネジ穴が写っており、その周囲は製品の平らで黒い表面に取り囲まれている。さらに外側にはカメラの囲いが写っている。ネジ穴データにおける異常及び ground truth を図 2 に示す。このような異常による品質の悪化は、製品の強度を低下させ、また破片は故障の原因となる。画像から読み取れるように、このデータセットにおいては、ネジ穴、平らな表面、ネジ穴の底面、などの複数の種類のクラスがあることがわかる。ネジ穴データセットは 12,406 枚の訓練データと、995 枚のテストデータから構成されている。テストデータの 995 枚の内、888 枚は正常データ、107 枚は異常データである。テストデータの正常、異常のラベル付けに関しては、アイシン・エイ・ダブリュ株式会社の専門家によって分類が行われており、これは手法の性能評価に用いられた。訓練データにはラベルが付いておらず、また訓練データはパラメータの調整に用いられた。訓練データも正常データと異常データが混在しており、各データの正確な割合は確認されていないが、テストデータと同程度の割合であるということをも前提として実験を行った。

4.2 モデル構成

提案手法を検証するにあたって、VAE、自己符号化器 (Autoencoders; AE) [Hinton 06]、GMM の 3 つのモデルを使用した。ここでは、それぞれのモデル構造及び学習方法について述べる。非正規化異常度の検証を行うにあたって、VAE による学習を行った。VAE の実装には PyTorch v0.1.12 を用いた。VAE の学習を行う際は、画像の一部をランダムに 96×96 のサイズで切り取り、このパッチを訓練データとした。テストを行う際は、同サイズの画像を 16 ピクセルの間隔で切り取り、このパッチをテストデータとした。このパッチ全てに関して異常度を算出し、少なくとも 1 枚が閾値を超えてい

る場合、テストデータは異常であると判断した。非正規化異常度 $\mathcal{L}_{VAE,M} = M_{VAE}(x)$ の評価のため、従来の異常度である $\mathcal{L}_{VAE} = D_{VAE}(x) + A_{VAE}(x) + M_{VAE}(x)$ を使用した異常検知も同様に行い、比較した。VAE の符号化器は 5 層の畳み込み層と 1 層の全結合層で構成した。n 番目の畳み込み層はカーネルサイズ 4×4 、ストライド 2 であり、 $N_c \times 2^{n-1}$ チャネルの特徴マップを出力し、出力にはさらに Batch Normalization [Ioffe 15] と活性化関数として ReLU [Nair 10] を適用した。全結合層からは平均ベクトル μ_z と標準偏差ベクトル σ_z を出力させた。この出力層は $2 \times N_z$ ユニットで構成されており、活性化関数として恒等関数を使用した。 N_z ユニットは平均ベクトル μ_z として使用し、それ以外は $\log \sigma_z^2$ の獲得のために使用した。復号化器の構造は、符号化器の構造と対になっており、符号化器と逆の構造となるようにした。復号化器の出力は、平均ベクトル μ_x と標準偏差ベクトル σ_x とした。VAE のパラメータ学習にあたっては勾配降下法を用いるが、この際の最適化アルゴリズムとしては Adam を用いた。Adam のパラメータは $\alpha = 10^{-3}$ 、 $\beta_1 = 0.9$ 、 $\beta_2 = 0.999$ とした。また荷重減衰を 0.9999 とした。畳み込み層の開始チャネル数 N_c は $N_c \in \{16, 32, 64\}$ から探索し、潜在空間 \mathcal{Z} における次元数 N_z は $\{2, 5, 10, 20, 50, 100\}$ から探索を行った。上記で述べた条件以外に関しては、VAE の元論文や、先行研究 [Ribeiro 17] に従った。

VAE の比較として、AE の評価も行った。AE は VAE の確定的な簡易モデルとしてみなすことができ、異常検知手法としても使用されている [Zhou 17]。AE は符号化器と復号化器から構成されており、符号化器による出力は潜在変数 z の点推定、復号化器による出力は入力データ x の再構成 \tilde{x} である。AE の目的関数は平均二乗誤差であり、これを異常度 $\mathcal{L}_{AE} = \frac{1}{N_x} \sum_{i=1}^{N_x} (x_i - \tilde{x}_i)^2$ として使用する。AE はモンテカルロサンプリングを必要としない。また、正則化項 $D_{VAE}(x)$ を持たず、標準偏差 σ_{x_i} は定数となる。AE の構造は VAE の構造と同一にしており、データセットの利用方法や勾配降下法における条件も同じである。畳み込み層の開始チャネル数 N_c は $N_c \in \{16, 32, 64\}$ から探索し、潜在空間 \mathcal{Z} における次元数 N_z は $\{2, 5, 10, 20, 50, 100\}$ から探索を行い、1 番優れたモデルの選択を行った。

GMM については、scikit-learn v0.19.1 における基本的なモデルを使用した。訓練データとして、1 枚あたり画像サイズ 96×96 のパッチを 100 枚ランダムにサンプルし、PCA を用いて各パッチから N_h 次元の特徴を抽出した。隠れクラスが各々の分散共分散行列を持つ (full covariance) GMM に対して、EM アルゴリズムを使用して抽出した特徴の学習を行う。PCA のコンポーネント数 N_h は $N_h \in \{2, 5, 10, 20, 50, 100, 200, 500\}$ から探索し、GMM の隠れクラスの数 N_z は $N_z \in \{2, 5, 10, 20, 50, 100\}$ から探索した。GMM の異常度として、負の対数尤度 \mathcal{L}_{GMM} と非正規化異常度 $\mathcal{L}_{GMM,M} = M_{GMM}(x)$ を用いて検証を行った。

4.3 異常検知の精度

評価指標として Receiver Operating Characteristic (ROC) 曲線の下部面積 (ROC-AUC) を用いた。ROC-AUC は最大値が 1.0 であり、値が 1.0 に近いほど異常検知性能が高いことを意味する。ROC 曲線及び ROC-AUC の結果を図 3 に示す。図 3 から、VAE における非正規化異常度の ROC-AUC が他の手法よりも高くなっていることがわかる。

次に、異常度の可視化を行う。異常度とは正常であるほど値が低く、正常とは違う状態であるほど値が高くなる。つまり、異常度を可視化することによって、その度合から異常箇所を発見することができるということになる。これを検証するために、異常箇所が予めわかっているデータに対して VAE にお

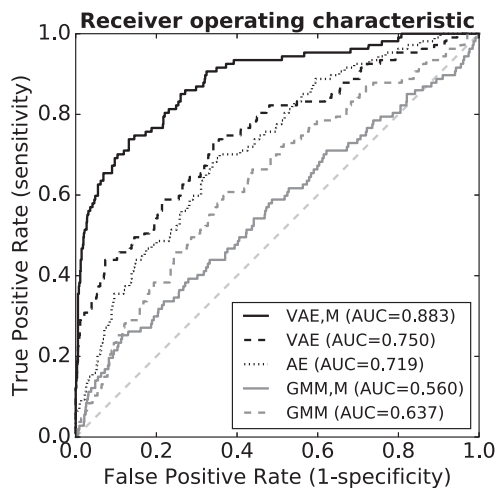


図 3: 各手法における ROC 曲線. 括弧内は ROC-AUC の値を表す.

る非正規化異常度 $\mathcal{L}_{VAE,M}$ を求め, その可視化を行うことで, 実際の異常箇所との比較を行う. 使用するモデルのパラメータは, ROC-AUC の 1 番高いものを採用する. 図 2 に示した異常データについて, VAE における従来の異常度 $\mathcal{L}_{VAE}(x)$ を用いたヒートマップ及び図非正規化異常度を用いたヒートマップを図 4 に示す. 異常度ヒートマップは, 画像を 96×96 のサイズに 4 ピクセルの間隔で切り取り, それぞれの異常度を並べて可視化したもので, 色が明るいほど異常度が低いことを示している. 異常のない部分の異常度について注目すると, 従来の VAE の異常度 $\mathcal{L}_{VAE}(x)$ はネジ穴の周囲の平らな黒い表面及びネジ穴の底面において小さな値を取っているが, ネジ穴の溝では正常にも関わらず大きな値を取っている. これはネジ穴の溝から得られたパッチは平らな部分から得られたパッチと比べて, 元の画像における相対的な位置の違いやネジ穴の溝の開始位置によって多様なものになり, また撮影方向や光の当たり具合によって変化するからであると考えられる. よって, ネジ穴の溝から得られたパッチの尤度は低くなる. つまり, 異常度が異常それ自体よりも, パッチ自体の複雑さに依存してしまっているということになる. 一方, 正規化距離 $M_{VAE}(x)$ はパッチが元画像において正常な部分から取られたものであれば, 場所によらず値が一定して低くなっている. 誤差の絶対値 $|x - \mu_x|$ はネジ穴の溝において大きい, 正規化距離 $M_{VAE}(x)$ は標準偏差 σ_x によって正規化されており, この標準偏差 σ_x は場所による複雑さを表しているため, 結果的に複雑さに対して堅牢となることで異常そのものを評価出来るようになると考えられる. このことから, 正規化距離 $M_{VAE}(x)$ は異常箇所においてのみ大きな値を取っている. 以上の理由により, 非正規化異常度 $\mathcal{L}_{VAE,M} = M_{VAE}(x)$ は複雑な構造をもつ工業製品の異常検知において, 性能の向上を達成する.

5. 結論

本論文では, VAE による異常検知において, 非正規化異常度を使用することを提案した. 非正規化異常度とは, その名前が指し示すように生成モデルの目的関数において正規化項を取り除いたものである. VAE において非正規化異常度を異常検知に使用することで, 複雑な構造を持つ工業製品の画像に対して堅牢となり, 従来の異常度よりも高い異常検知性能を実現

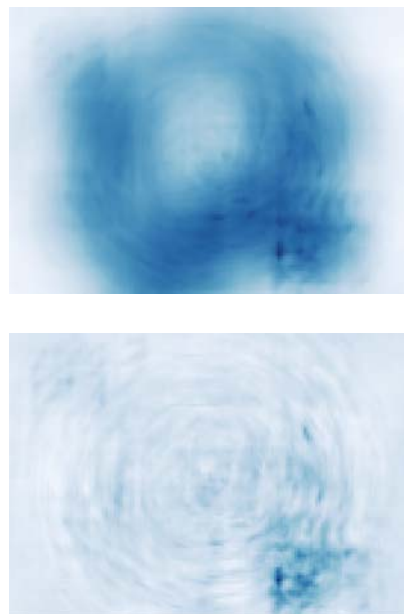


図 4: (上図) 従来の異常度 $\mathcal{L}_{VAE}(x)$ のヒートマップ. (下図) 非正規化異常度 $\mathcal{L}_{VAE,M}$ のヒートマップ.

することが出来た. また, 異常度の可視化を行うことにより, 異常箇所の特定を行えることを示した.

本研究は科研費 (16K12487) とアイシン・エイ・ダブリュ株式会社, 株式会社エイ・ダブリュ・エンジニアリングの支援, 総務省 SCOPE(受付番号 172107101) の委託を受けて行われた.

参考文献

- [Hinton 06] Hinton, G. E. and Salakhutdinov, R. R.: Reducing the Dimensionality of Data with Neural Networks, *Science*, Vol. 313, No. 5786, pp. 504–507 (2006)
- [Ioffe 15] Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in *International Conference on Machine Learning (ICML)* (2015)
- [Nair 10] Nair, V. and Hinton, G. E.: Rectified Linear Units Improve Restricted Boltzmann Machines, in *International Conference on Machine Learning (ICML)*, pp. 807–814 (2010)
- [Ribeiro 17] Ribeiro, M., Lazzaretti, A. E., and Lopes, H. S.: A study of deep convolutional autoencoders for anomaly detection in videos, *Pattern Recognition Letters* (2017)
- [Saligrama 12] Saligrama, V. and Chen, Z.: Video anomaly detection based on local statistical aggregates, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2112–2119 (2012)
- [Zhou 17] Zhou, C. and Paffenroth, R. C.: Anomaly Detection with Robust Deep Autoencoders, *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 665–674 (2017)