サマースクール：画像認識 第4回 異常検知

北田敦也

2021/09/07

今回の目的・目標

目標

- 異常の特性や異常検知特有の問題設定を踏まえて、画像に関する各種異常検知手法とその特性を理解し、一部の手法を実装できるようにする

目的

- 異常の性質を理解し、身の回りの異常を適切に分類・説明できる
- 教師なしの異常検知に関して、クラス分類手法、確率推定手法、再構成手法のそれぞれについて従来手法と比較しながらモデルの説明ができる
- 欠陥製品画像から異常を検知するプログラムを実装できる

目次

導入

異常とは

身の回りにある異常

- 病気・怪我、交通事故、故障・欠陥、システム侵入、盗難・強盗、テロ、不正取引.....

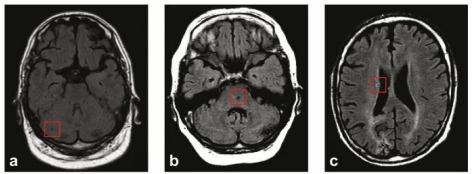
異常とは

言葉による定義

- “An anomaly is an observation that deviates considerably from some concept of normality”

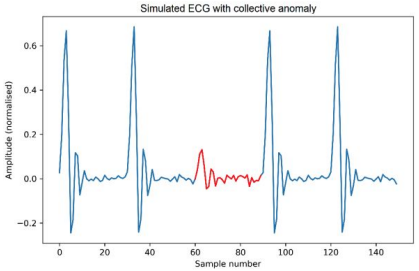
  「異常とは、正常という概念から大きく逸脱した観測結果」（Ruff, 2021 [1]）

医療画像



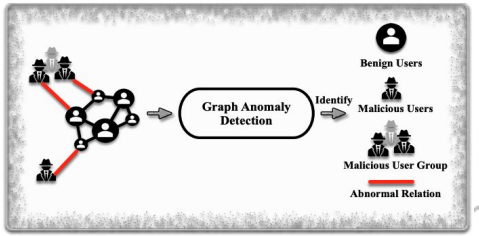
[2] Hespen, Kees M. van, Jaco J. M. Zwanenburg, Jan W. Dankbaar, Mirjam I. Geerlings, Jeroen Hendrikse, and Hugo J. Kuijf. “An Anomaly Detection Approach to Identify Chronic Brain Infarcts on MRI.” *Scientific Reports* 11, no. 1 (April 8, 2021): 7714. より図を引用

心電図



[3] Cook, Andrew A., Göksel Mısırlı, and Zhong Fan. “Anomaly Detection for IoT Time-Series Data: A Survey.” IEEE Internet of Things Journal 7, no. 7 (July 2020): 6481–94. より図を引用

SNSの不正アカウント



[4] Ma, Xiaoxiao, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z. Sheng, Hui Xiong, and Leman Akoglu. “A Comprehensive Survey on Graph Anomaly Detection with Deep Learning.” ArXiv:2106.07178 [Cs], August 24, 2021. より図を引用

[1] Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. “A Unifying Review of Deep and Shallow Anomaly Detection.” *Proceedings of the IEEE* 109, no.5 (May 2021): 756–95. より引用

用語の整理

異常のさまざまな定義

Anomaly

- 正常な分布とは別の分布から生じるサンプル

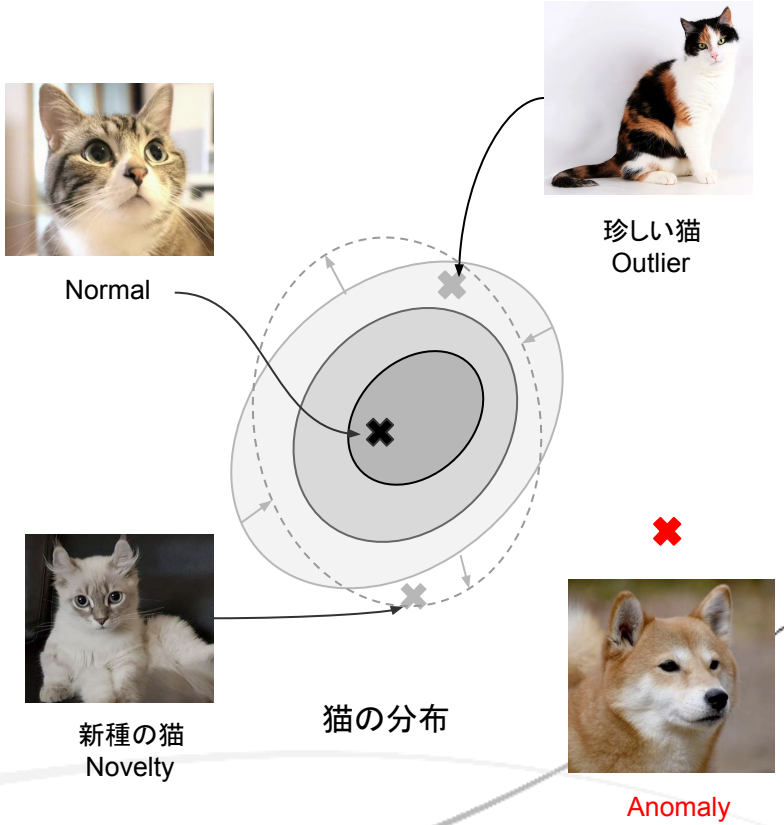
Outlier

- 正常な分布から低確率で生じるサンプル

Novelty

- 正常な分布の変化などにより新たに観測されたサンプル

※ 検出方法はどれも似ているため、ここでは区別しない



[1] Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. "A Unifying Review of Deep and Shallow Anomaly Detection." *Proceedings of the IEEE* 109, no.5 (May 2021): 756–95. より定義を引用

用語の整理

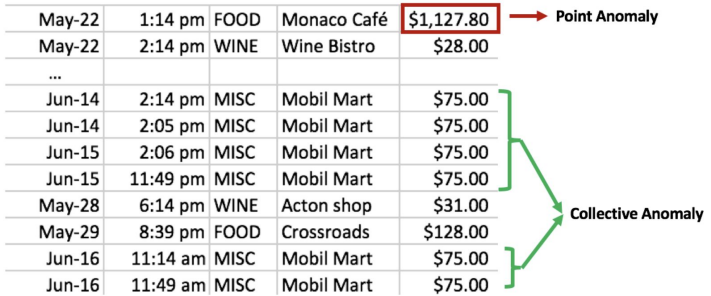
異常のタイプ

Point Anomaly

- 一つのデータが他のデータと比較して異常な場合
- 最も単純で、一般的なタイプの異常
- 例：クレジットカードの高額決済

Group Anomaly

- 個々のデータは一見正常だが、全体として異常な性質を示すもの
- ex) クレジットカードの連続決済、SNSでの大量スパムアカウント



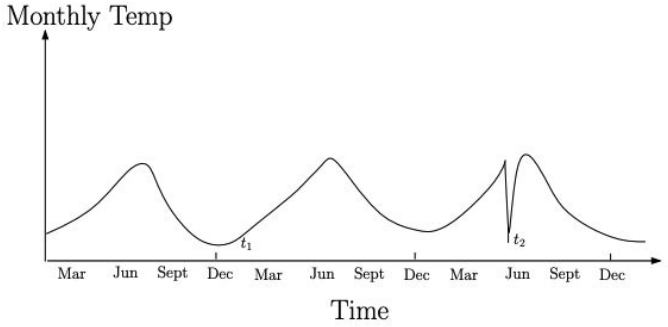
[5] Chalapathy, Raghavendra, and Sanjay Chawla. "Deep Learning for Anomaly Detection: A Survey." *ArXiv:1901.03407 [Cs, Stat]*, January 23, 2019. より図を引用

用語の整理

異常のタイプ

Contextual Anomaly

- 特定の条件下でのみ異常となるデータ
- 時間・場所・行動パターンといった周囲の条件により決まる
- 時系列データに多い
- ex) 冬+気温10℃→正常、夏+気温10℃→異常



[6] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly Detection: A Survey." *ACM Comput. Surv.* 41 (July 2009). より図を引用

具体的事例

侵入検知

- コンピュータシステムへの侵入や不正使用
- システムコールの系列データ
- Group Anomaly
  - 複数のシステムコールをまとめて考慮する必要がある

open,	read,	mmap,	mmap,	open,	read,	mmap	...
open,	mmap,	mmap,	read,	open,	close	...	
open,	close,	open,	close,	open,	mmap,	close	...

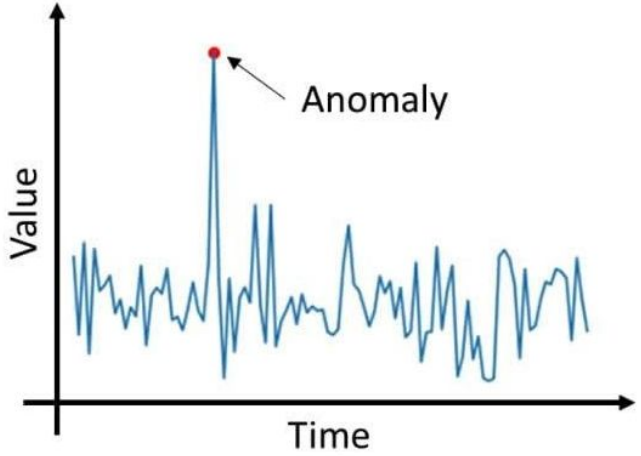
システムコールの記録

[6] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. “Anomaly Detection: A Survey.” *ACM Comput. Surv.* 41 (July 2009). より図を引用

具体的事例

機械の故障・欠陥

- モーター、エンジンなどで発生する故障
- センサーによって収集した音声などの時系列データ
- Contextual Anomaly
  - 周囲の値の変化と比較して異常を判断する
- Group Anomaly
  - 異常値が続けて発生した場合

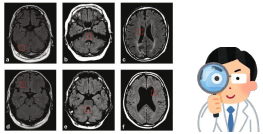


[7]https://www.countants.com/blogs/how-machine-learning-can-enable-anomaly-detection/
より図を引用

異常検知特有の問題設定

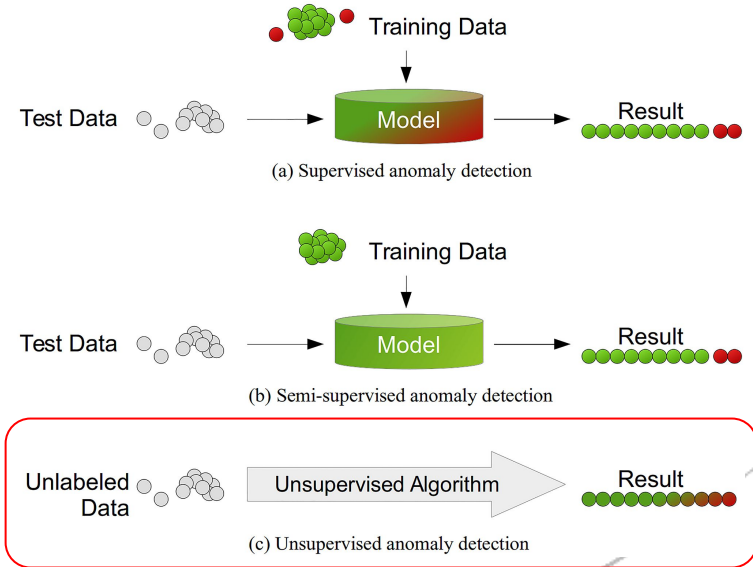
異常データの入手困難性

- 異常・正常が不均衡
  - 正常データは大量
  - 異常データは稀にしか発生せず、極端に少量
- 専門知識が必要で、異常データのアノテーションコストが高い



- **教師なし学習** >>> 教師あり学習

→ **本講義では教師なし学習手法にフォーカス**

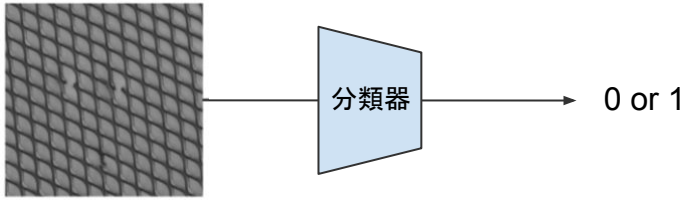


[8] Goldstein, Markus, and Seiichi Uchida. "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data." *PLOS ONE* 11, no. 4 (April 19, 2016): e0152173.より図を引用

画像分野でのタスク

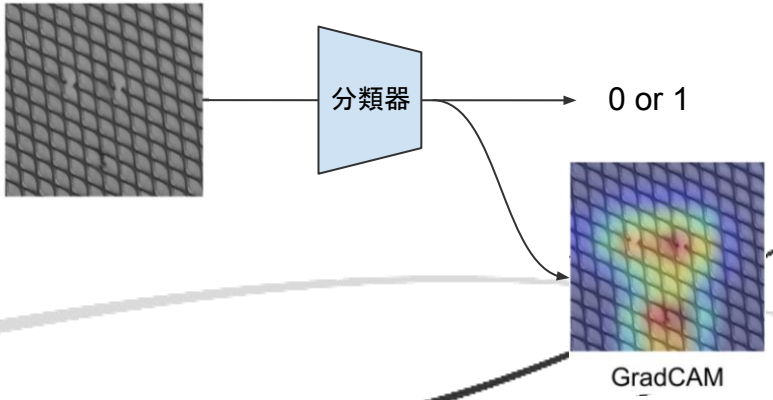
異常・正常の分類（本講義で扱うタスク）

- 教師なし学習がメインで行うタスク
- 欠陥製品画像の分類など



異常箇所の特定

- 教師なしでは難易度高い
- 病気、欠陥部位の検出など
- Grad-CAM等の利用



教師なし深層異常検知の流れ

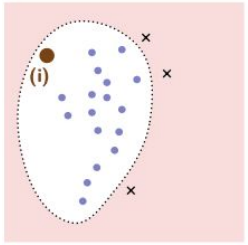
“An anomaly is an observation that deviates considerably from some concept of normality”

「異常とは、正常という概念から大きく逸脱した観測結果」

1. 正常データの特徴（データの存在範囲、確率分布など）をうまく表現するようにモデルを学習
2. 異常度合いを表すスコアリング関数s(x)を定義
3. 異常スコアの閾値 τ を定め、それを超えたものを異常と判定

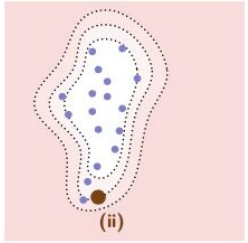
$$A = \{x \in X \mid s(x) \geq \tau\}$$

手法の分類



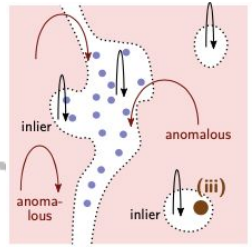
1. One-Class Classification Model

- 正常データが属するクラスの決定境界を引き、その外部にあるデータを異常とみなす



2. Probabilistic Model

- 正常データの確率分布を推定し、低確率領域にあるデータを異常とみなす

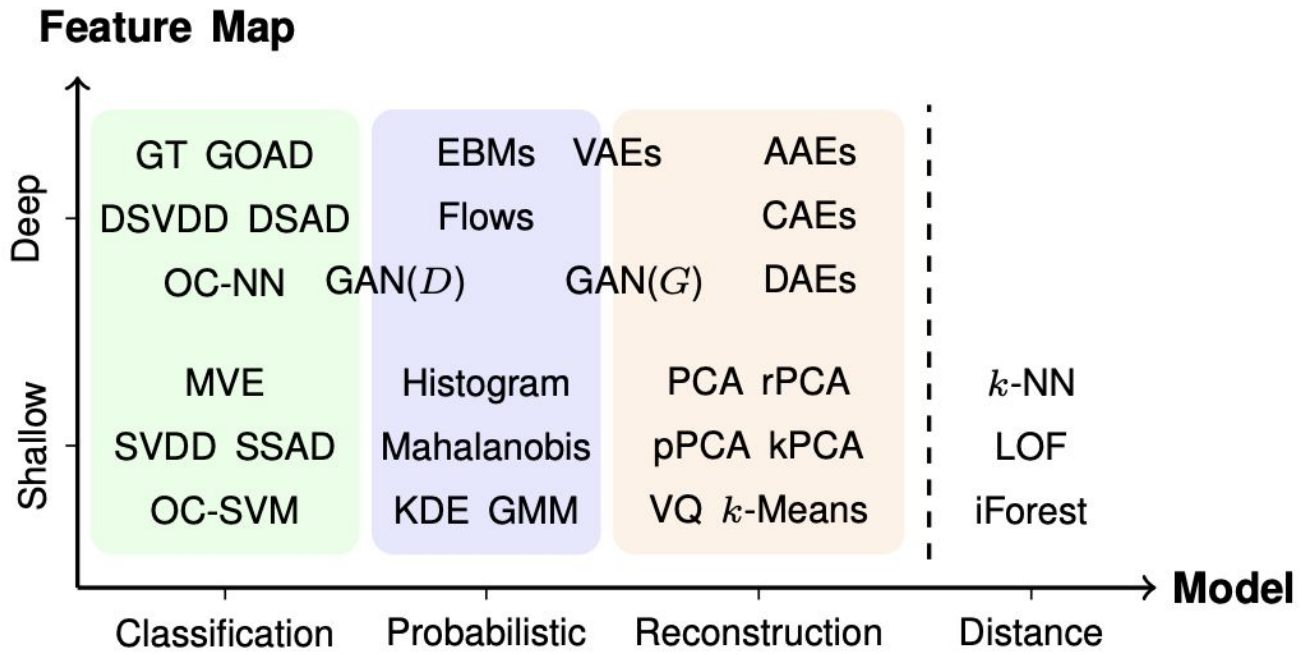


3. Reconstruction Model

- 正常データを復元するようにモデルを学習し、うまく復元できないデータを異常とみなす

[1] Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. "A Unifying Review of Deep and Shallow Anomaly Detection." *Proceedings of the IEEE* 109, no.5 (May 2021): 756–95. より図を引用

手法の分類

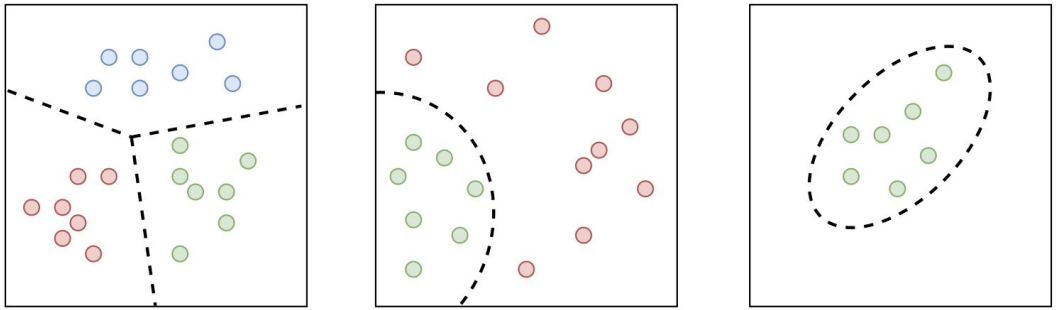


[1] Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. "A Unifying Review of Deep and Shallow Anomaly Detection." *Proceedings of the IEEE* 109, no.5 (May 2021): 756–95. より図を引用

One-Class Classification Model

One-Class Classification Model の概要

- One Class Classificationとは、1クラスのデータしか手に入らない場合の2値分類タスク
- 正常データがほどよく収まるような決定境界を学習する
  - 見逃し（FN）を最小に抑えるためにできるだけ多くの正常データ点を含みつつ、誤検出（FP）を抑えるためにできるだけ範囲を狭くする



[9] erera, Pramuditha, Poojan Oza, and Vishal M. Patel. “One-Class Classification: A Survey.” *ArXiv:2101.03064 [Cs]*, January 8, 2021. より図を引用

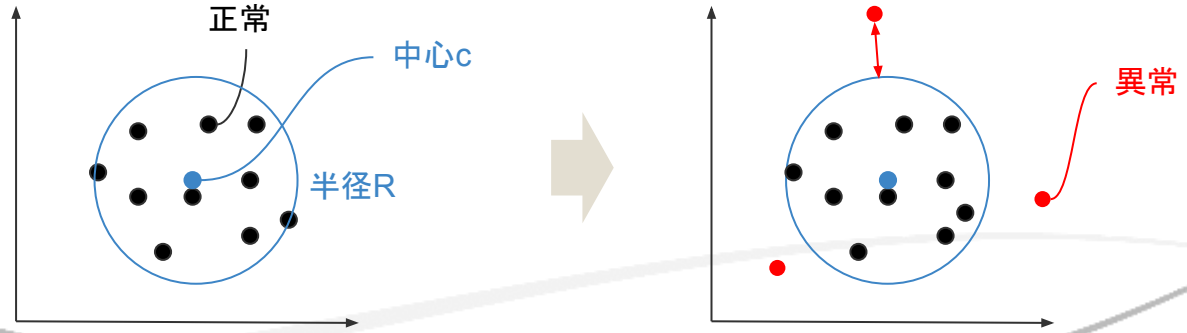
One-Class Classification Model – 従来手法 –

SVDD (Support Vector Data Discription)

- 異常と正常を分離する境界を決定する
  - 正常データのほぼ全てを含みつつ、できる限り小さい超球になるように半径を最適化
- 最適化した超球との距離で異常度を測る

$$\min_{R,\boldsymbol{c},\boldsymbol{\xi}} \quad R^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i$$

$$\text{s.t.} \quad \|\boldsymbol{x}_i - \boldsymbol{c}\|^2 \le R^2 + \xi_i, \quad \xi_i \ge 0, \quad \forall i.$$



[10] D. M. J. Tax and R. P. W. Duin, “Support vector data description,” Machine Learning, vol. 54, no. 1, pp. 45–66, 2004.

One-Class Classification Model – 深層学習手法 –

Deep SVDD

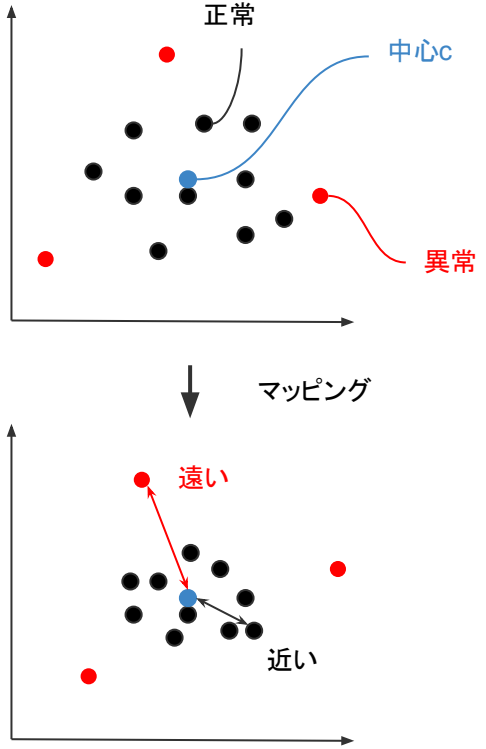
- 全データの中心点cへ向かって正常データを写像するようにニューラルネットワークΦ(x; W)を学習する。

$$\min_{\mathcal{W}} \quad \frac{1}{n}\sum_{i=1}^{n}\|\phi(\boldsymbol{x}_i;\mathcal{W})-\boldsymbol{c}\|^2+\frac{\lambda}{2}\sum_{\ell=1}^{L}\|\boldsymbol{W}^{\ell}\|_F^2$$

中心に近いほど小さい損失に

- マッピングしたデータΦ(x)と中心cからの距離を、そのまま異常スコアとして使用
  - 正常データであれば中心cの近くに、異常データであれば中心cから遠くにマッピングされる

$$s(\boldsymbol{x})=\|\phi(\boldsymbol{x};\mathcal{W}^*)-\boldsymbol{c}\|^2$$



[11] Ruff, Lukas, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. “Deep One-Class Classification.” In International Conference on Machine Learning, 4393–4402. PMLR, 2018.

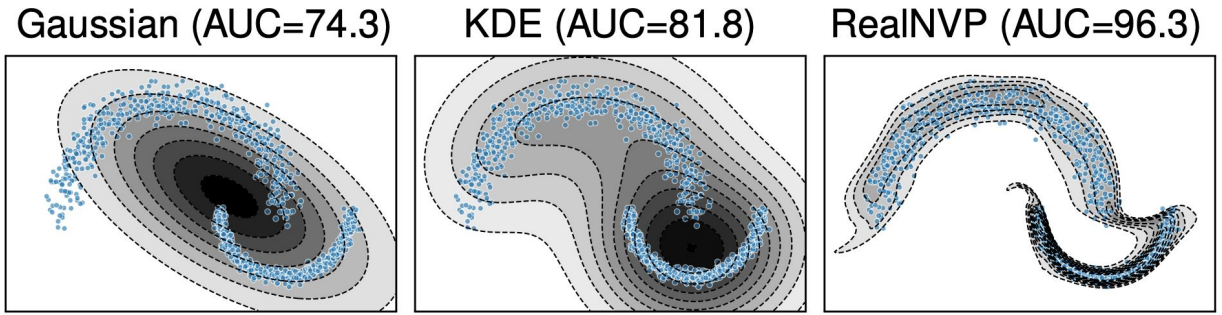
Probabilistic Model

Probabilistic Model の概要

- 正常データの確率分布をモデルによって近似

$$p_\theta(x) \approx p^+(x)$$

- 確率を異常スコアとし、低確率領域に存在するデータを異常とみなす
- 確率モデルの表現方法によって様々な手法が存在



[1] Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. "A Unifying Review of Deep and Shallow Anomaly Detection." *Proceedings of the IEEE* 109, no.5 (May 2021): 756–95. より図を引用

Probabilistic Model – 従来手法 –

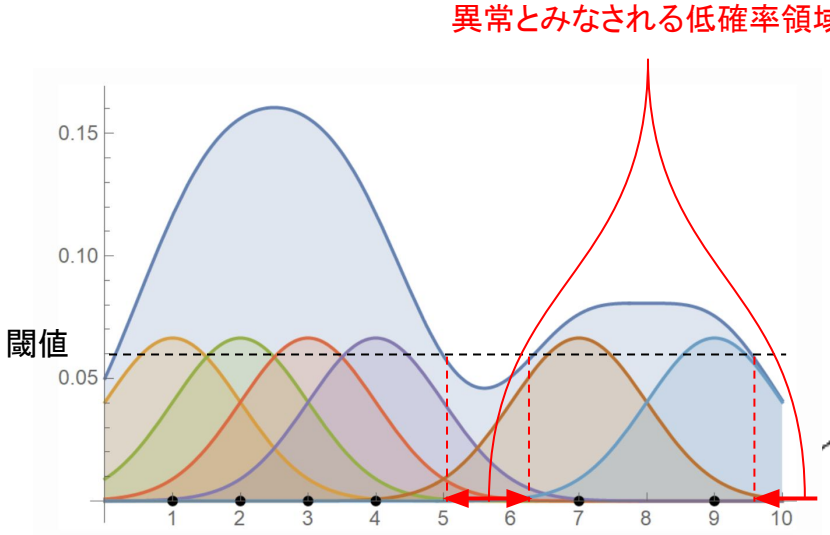
KDE (Kernel Density Estimation)

- 各データ点x周りのカーネル関数K(x)（通常は正規分布を利用）を重ね合わせていくことで、全体のデータ分布p(x)を表現

$$\hat{p}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

$$K(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{x^2}{2}\right)$$

- 低確率領域にあるデータを異常と判定

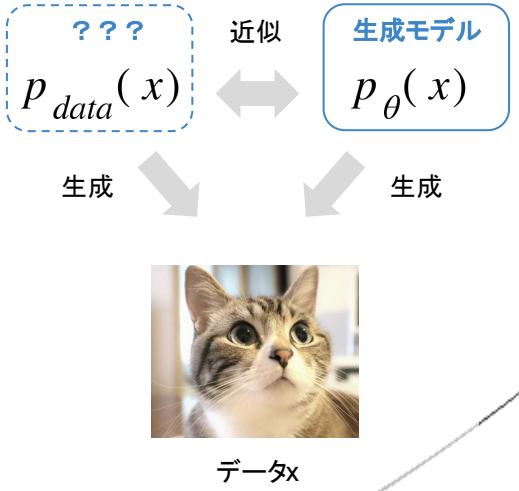


[12] https://ekamperi.github.io/math/2020/12/08/kernel-density-estimation.html より図を引用

Probabilistic Model – 深層学習手法 –

生成モデルの復習

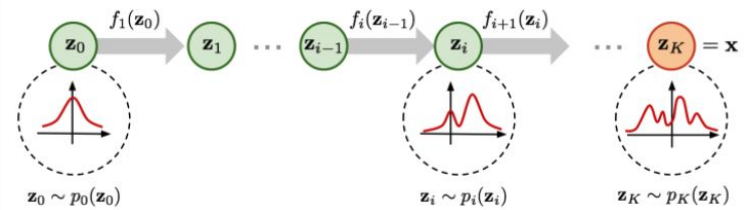
- データxはある確率分布p_data(x)から独立に生成されていると仮定
  - あくまで仮定であり、実際にそのような確率分布は手元になく形も不明
- データ分布p_data(x)は手に入らないので、代わりにデータxの確率モデルp_θ(x)を用意
  - モデル構造は設計者による仮定（例：二項分布、正規分布など）
  - θは確率モデルのパラメータ（例：正規分布における平均、標準偏差）
- パラメータθを調整することで確率モデルp_θ(x)によって、データ分布p_data(x)を近似
  - 確率モデルから実際のデータに似たデータが生成されるようになる
  - この時のp_θ(x)をxの **生成モデル** と呼ぶ

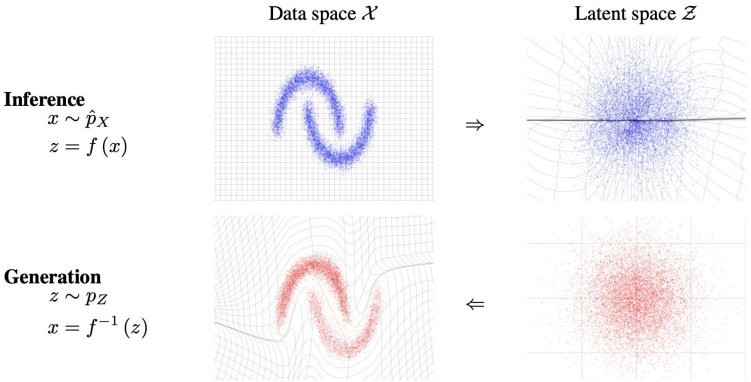


Probabilistic Model – 深層学習手法 –

フローベース

- 潜在変数を単純な事前分布p(z)（ガウシアン等）から生成し、それに可逆な変換を繰り返し適用して目標の分布p(x)からのサンプルを得る
- データの生成過程が可逆な変換のため、データから潜在変数への正確な変換が可能
- 複雑な分布に従うデータxを、単純な分布に従う潜在変数zに変換し、データxの生じる確率を算出





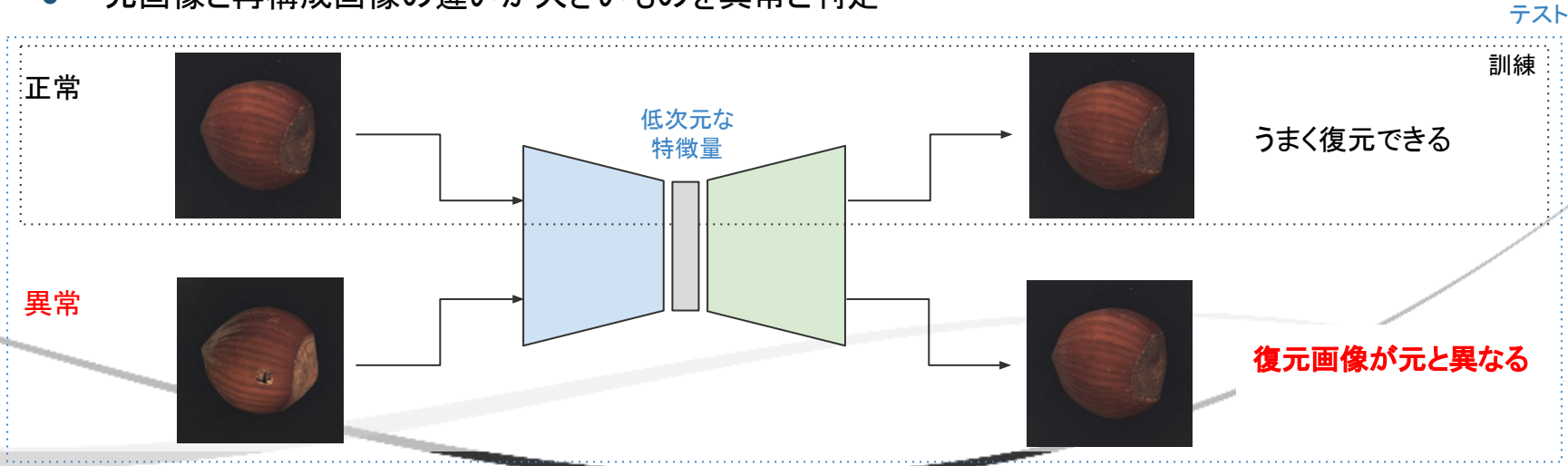
[13] Dinh, Laurent, Jascha Sohl-Dickstein, and Samy Bengio. “Density Estimation Using Real NVP.” *ArXiv:1605.08803 [Cs, Stat]*, February 27, 2017. より図を引用

Reconstruction Model

Reconstruction Modelの概要

異常データは上手く再構成されない

- 正常データを低次元な特徴量へ圧縮して元の正常データへと再構成するモデルを学習
- 異常データは正常データと異なる特徴を持つため、正確に再構成されない
- 元画像と再構成画像の違いが大きいものを異常と判定

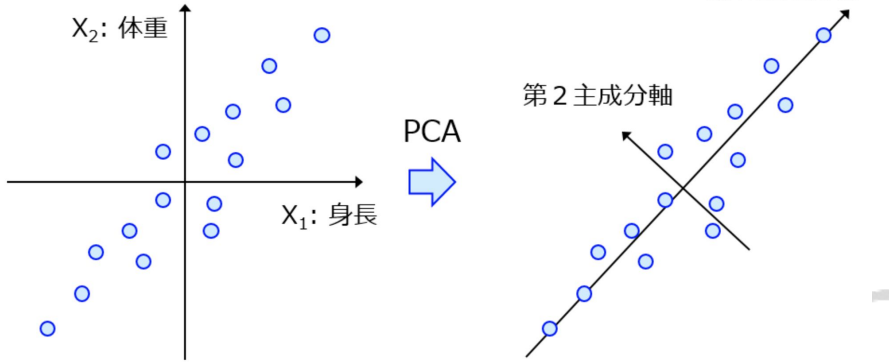


[14] https://www.mvtec.com/company/research/datasets/mvtec-adより図を引用

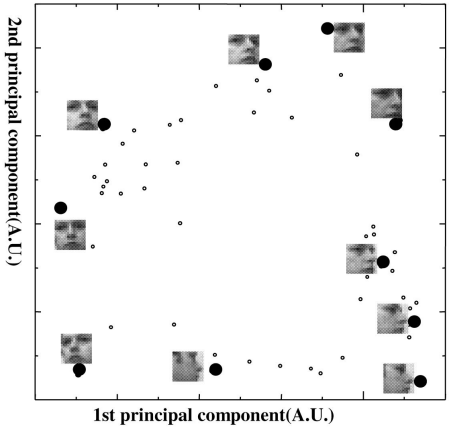
Reconstruction Model – 従来手法 –

PCA（主成分分析）

- データのばらつきが最大になるような新たな軸を見つける
- さらにその軸と直交し、データのばらつきが最大になるような軸を見つける
- これを繰り返し、元の次元より少ない次元数に圧縮する
- 圧縮した点をさらに元の空間に逆写像したデータと元データが離れていれば異常とみなす



[15] https://datachemeng.com/principalcomponentanalysis/より図を引用

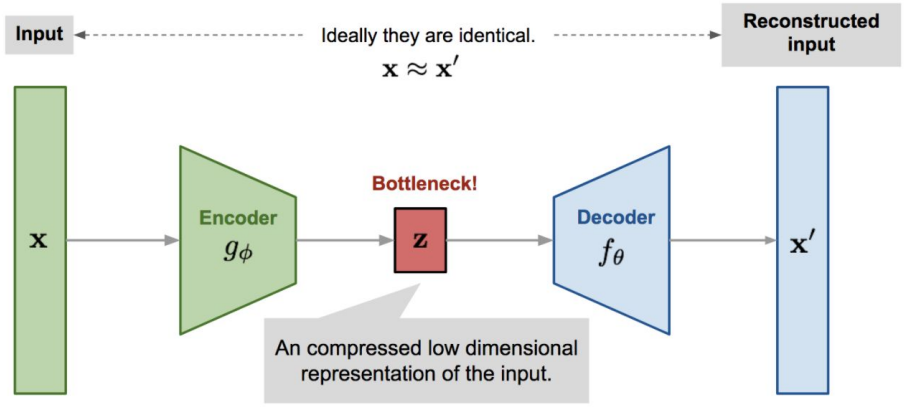


[16] 坂野「パターン認識における主成分分析　--顔画像認識を例として（研究詳解）(特集 地図を描く・風景を眺める --主成分分析・多次元尺度法とその周辺　)」統計数理 (2001) 第 49 巻 第 1 号 23–42 より図を引用

Reconstruction Model – 深層学習手法 –

Auto Encoder

- Encoderにより入力画像を低次元な特徴量へと圧縮し、Decoderによって元の画像へ再構成
- 入力画像と生成画像の誤差が大きい画像を異常とみなす
- 画像がぼやけやすいのが難点



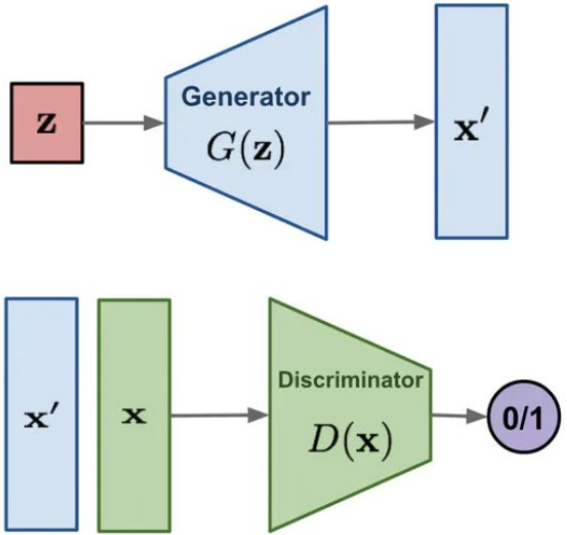
[17] https://lilianweng.github.io/lil-log/2018/08/12/from-autoencoder-to-beta-vae.htmlより図を引用

Reconstruction Model – 深層学習手法 –

GANの復習

- 生成器 Generator：潜在変数zから偽物の画像を生成
- 識別器 Discriminator：本物画像xと偽物画像G(z)がランダムに入力され、本物か偽物かを識別
- 学習：GはDを騙すように、DはGに騙されないように、GとDを交互に最適化

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

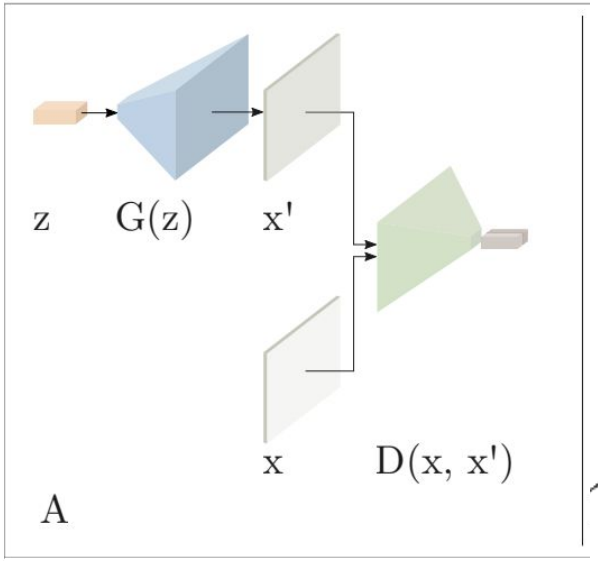


[17]https://lilianweng.github.io/lil-log/2018/10/13/flow-based-deep-generative-models.htmlより図を引用

Reconstruction Model – 深層学習手法 –

AnoGAN 1/3

- 初めて異常検知にGANを応用したモデル
- 3ステップに分かれる
  a. 正常データでGANの学習
  b. テスト入力画像xに対応する潜在変数zを見つける
  c. 入力画像xとそれに対応する生成画像G(z)の差分から異常を検出
- 正常画像のみでGANを学習



[18] Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training." *ArXiv:1805.06725 [Cs]*, November 13, 2018. より図を引用

[19] Schlegl, Thomas, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery." ArXiv:1703.05921 [Cs], March 17, 2017.

Reconstruction Model – 深層学習手法 –

AnoGAN 2/3

- 入力画像xに対応する潜在変数zの探索
  - Residual Loss：入力画像と生成画像のピクセルごとの差

$$\ell_R(\mathbf{x}, \mathbf{z}_\gamma) = \left\|\mathbf{x} - G(\mathbf{z}_\gamma)\right\|$$

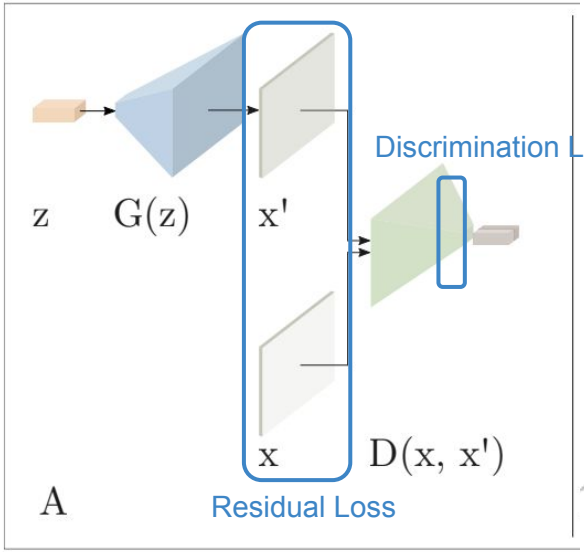
  - Discrimination Loss：Dの最終層手前の特徴量における、ピクセルごとの差

$$\ell_{fm}(\mathbf{x}, \mathbf{z}_\gamma) = \left\|h(\mathbf{x}) - h\left(G(\mathbf{z}_\gamma)\right)\right\|$$

  - 2つのLossの合計が小さくなるようなzを求める

$$s_\mathbf{x} = (1-\alpha)\ell_R(\mathbf{x}, \mathbf{z}_{\gamma^*}) + \alpha\ell_{fm}(\mathbf{x}, \mathbf{z}_{\gamma^*})$$

α：2種類のLossのバランスをとる



[18] Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. “GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training.” *ArXiv:1805.06725 [Cs]*, November 13, 2018. より図を引用

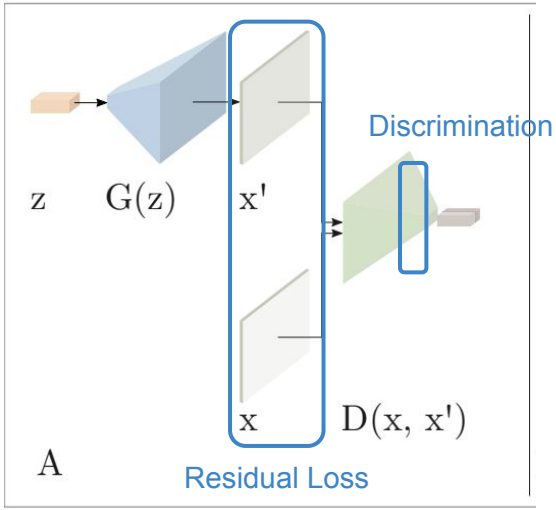
Reconstruction Model – 深層学習手法 –

AnoGAN 3/3

- 入力画像xとそれに対応する生成画像G(z)の差分から異常を検出
- 異常スコアにはzの最適化に用いたLossを用いる

$$\ell_R(\mathbf{x}, \mathbf{z}_\gamma) = \left\|\mathbf{x} - G(\mathbf{z}_\gamma)\right\|$$
$$\ell_{fm}(\mathbf{x}, \mathbf{z}_\gamma) = \left\|h(\mathbf{x}) - h\left(G(\mathbf{z}_\gamma)\right)\right\|$$
$$s_\mathbf{x} = (1-\alpha)\ell_R(\mathbf{x}, \mathbf{z}_{\gamma^*}) + \alpha\ell_{fm}(\mathbf{x}, \mathbf{z}_{\gamma^*})$$

- zの探索に時間がかかるという欠点



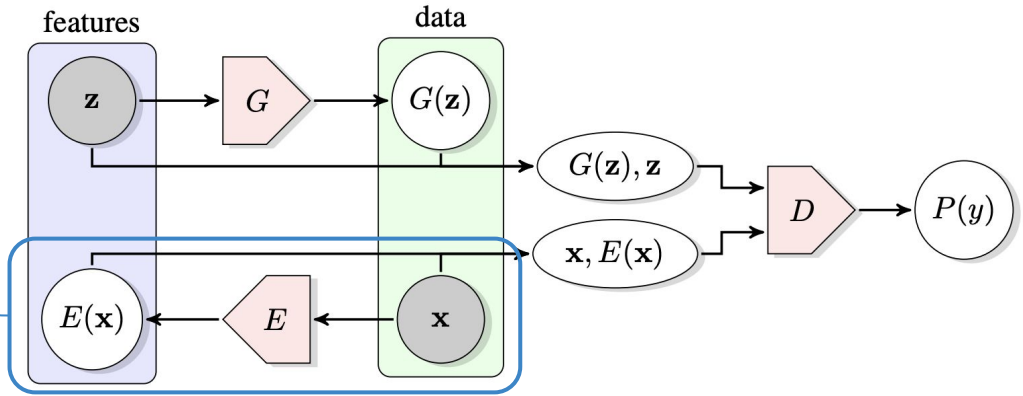
[18] Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training." *ArXiv:1805.06725 [Cs]*, November 13, 2018. より図を引用

Reconstruction Model – 深層学習手法 –

Efficient-GAN 1/3

- BiGANを利用
  - GとDの学習と同時に、入力画像xを潜在変数zに変換するエンコーダーEも学習
  - 識別器は画像と潜在変数のセットを入力として受け取り、真偽判定する
- 推論時に画像xに対応する潜在変数zを見つける手間を省く

画像xを潜在変数zに変換するエンコーダー



[20] Donahue, Jeff, Philipp Krähenbühl, and Trevor Darrell. “Adversarial Feature Learning.” ArXiv:1605.09782 [Cs, Stat], April 3, 2017. より図を引用

[21] Zenati, Houssam, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Ramaseshan Chandrasekhar. “Efficient GAN-Based Anomaly Detection.” ArXiv:1802.06222 [Cs, Stat], May 1, 2019.

Reconstruction Model – 深層学習手法 –

Efficient-GAN 2/3

- 学習時にはDとGに加えてエンコーダーEも最適化
- 識別器Dは、本物画像のペア(x, E(x))と生成画像のペア(G(z), z)を識別したい
  - V(D, E, G)の第1項と第2項を最大化
- 生成器Gは、xとG(z)が識別されないようにしたい
  - V(D, E, G)の第2項を最小化（第1項はGが関与しない）
- エンコーダーEは、E(x)とzが区別されないようにしたい
  - V(D, E, G)の第1項を最小化（第2項はEが関与しない）

$$\min_{G,E} \max_{D} V(D,E,G)$$

$$V(D,E,G) = \mathbb{E}_{x\sim p_X}\left[\mathbb{E}_{z\sim p_E(\cdot|x)}\left[\log D(x,z)\right]\right] + \mathbb{E}_{z\sim p_Z}\left[\mathbb{E}_{x\sim p_G(\cdot|z)}\left[1-\log D(x,z)\right]\right]$$

E(x)

G(z)

本物画像を本物とみなすと1を取る

生成画像を偽物とみなすと1を取る

Reconstruction Model – 深層学習手法 –

Efficient-GAN 3/3

- 異常スコアの算出は基本的にAnoGANと同じ
- 識別器に画像だけでなく潜在変数も入力するため、Discrimination Lossの形がやや違う

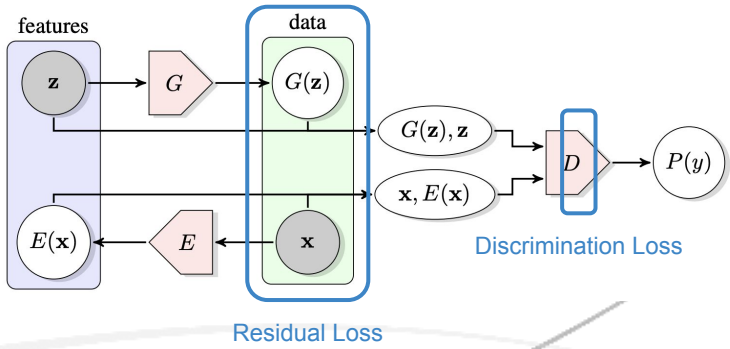
$$A(x) = \alpha L_G(x) + (1-\alpha) L_D(x)$$

$$L_G(x) = ||x - G(E(x))||$$

生成画像とテスト画像のピクセルごとの差

$$L_D(x) = ||f_D(x, E(x)) - f_D(G(E(x)), E(x))||$$

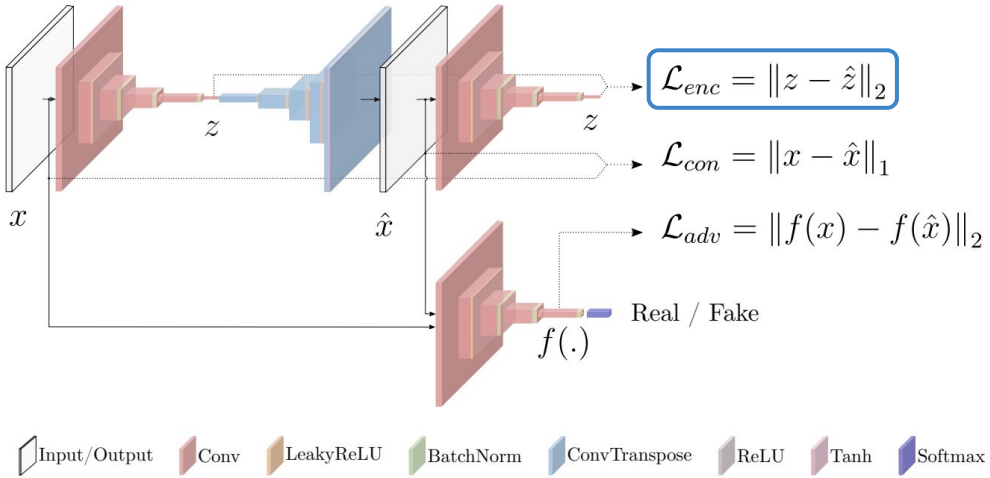
Dの最終全結合層の手前の特徴量における、ピクセルごとの差



Reconstruction Model – 深層学習手法 –

GANomaly

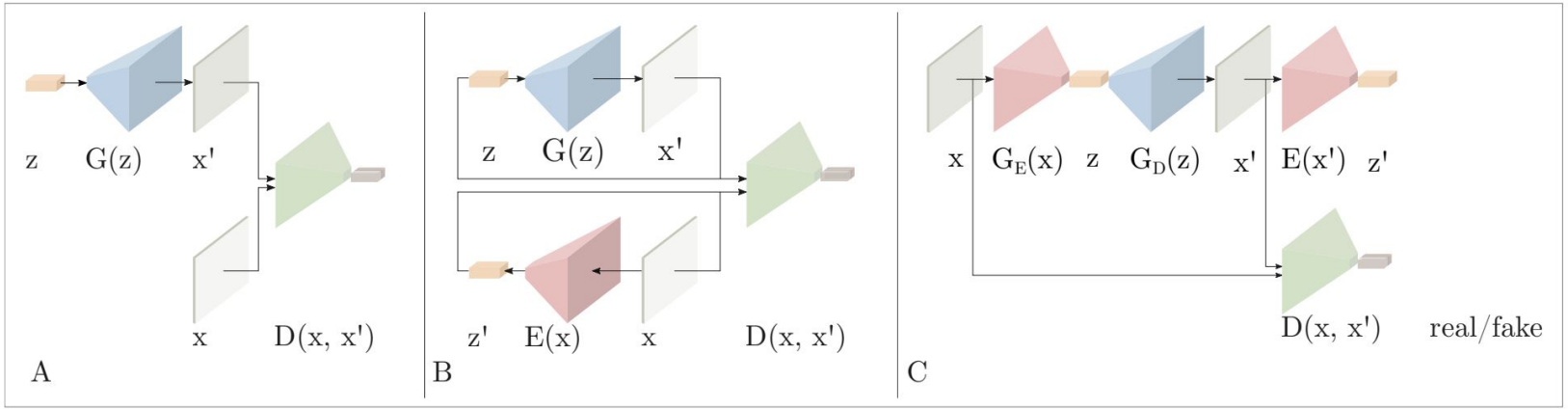
- GANとAEの組み合わせ
- 新たに、入力画像xに対応する潜在変数zと、生成画像G(x)に対応する潜在変数zをLossで比較
- Efficient-GANと異なり、画像のみを識別器Dに入力



[18] Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. "GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training." *ArXiv:1805.06725 [Cs]*, November 13, 2018. より図を引用

Reconstruction Model – 深層学習手法 –

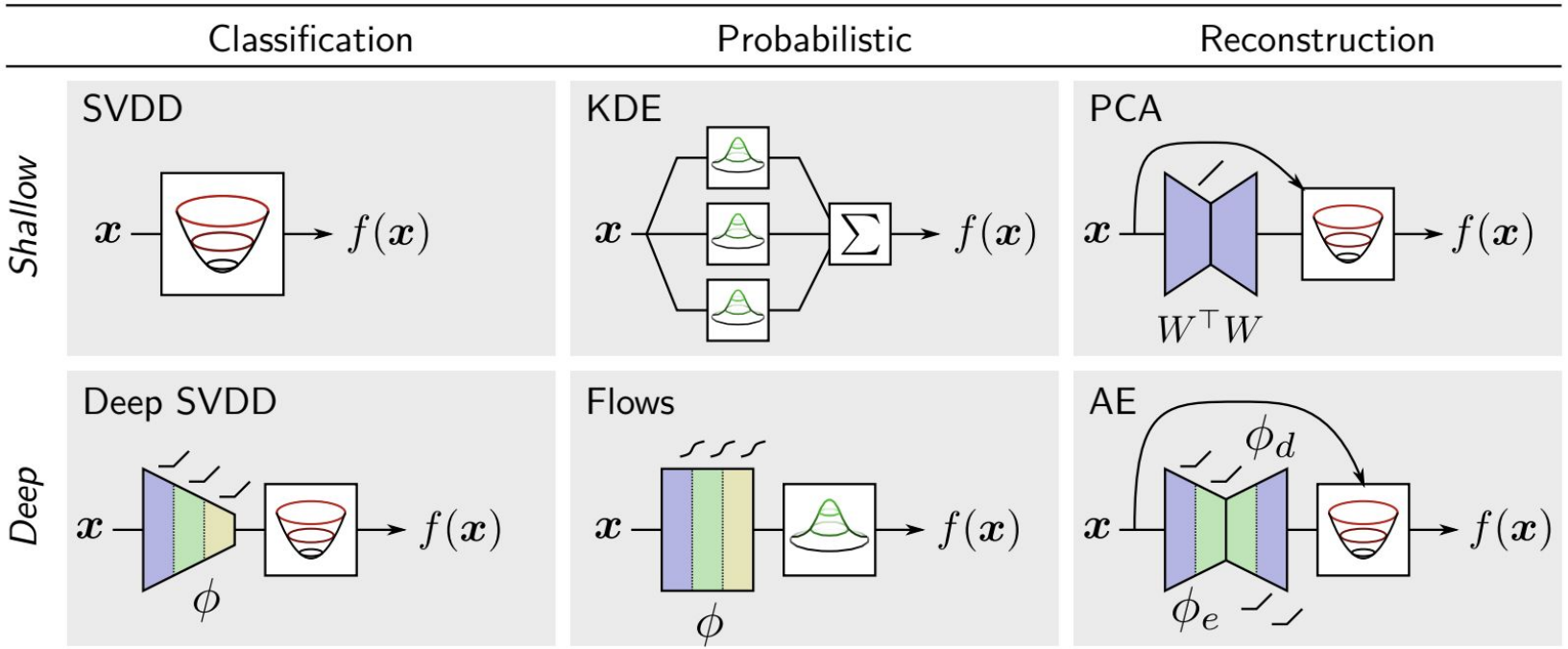
- AnoGANはGANの機構をそのまま取り入れた最初のモデル
- Efficient-GANやGANomalyではEncoderを一緒に学習させることで、入力画像に対応する潜在変数zの探索に時間を要するというAnoGANの問題点を解決



AnoGAN　　Efficient-GAN　　GANomaly

[18] Akcay, Samet, Amir Atapour-Abarghouei, and Toby P. Breckon. “GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training.” *ArXiv:1805.06725 [Cs]*, November 13, 2018. より図を引用

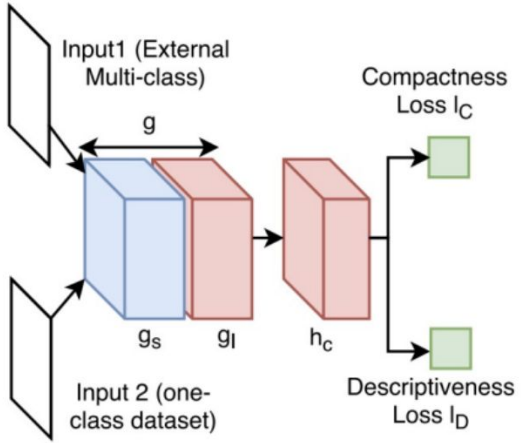
各手法のまとめ

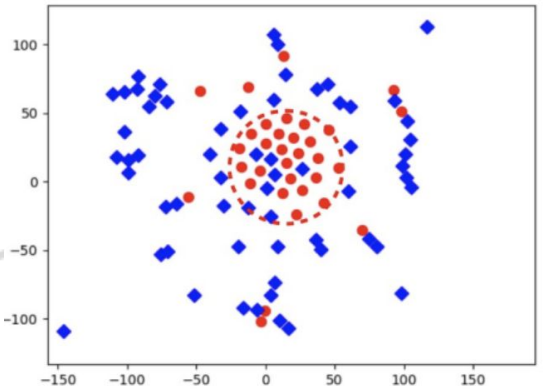


[1] Ruff, Lukas, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. "A Unifying Review of Deep and Shallow Anomaly Detection." *Proceedings of the IEEE* 109, no.5 (May 2021): 756–95. より図を引用

その他手法 - 距離学習 -

- 同じクラスは近くに、異なるクラスは遠くになるように学習
- 基本的に多クラス分類タスクで利用されるが、異常検知では1クラスしか手に入らない（正常クラス）
- 正常データのみからなるターゲットデータセットと、複数クラスからなるリファレンスデータセットを用意
- リファレンスデータセットでは分類問題を、ターゲットデータセットでは特徴量が平均に近づくように学習

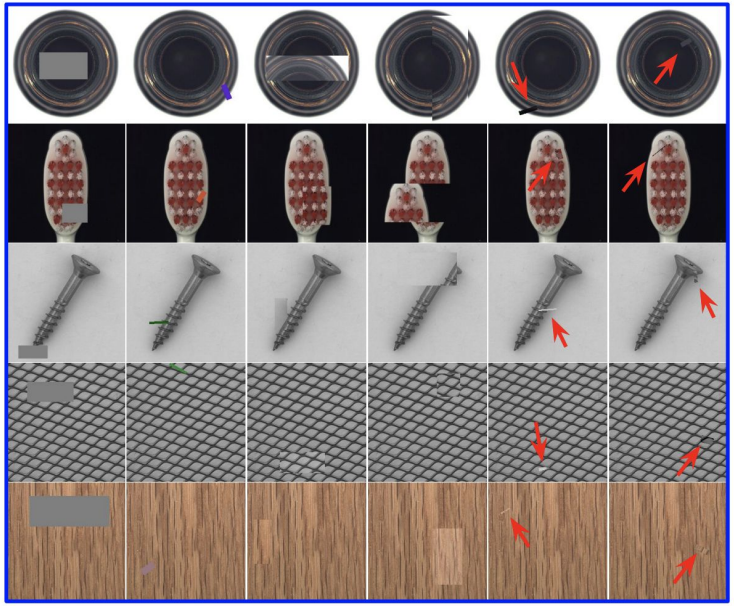




[22] Perera, Pramuditha, and Vishal M. Patel. “Learning Deep Features for One-Class Classification.” IEEE Transactions on Image Processing 28, no. 11 (November 2019): 5450–63. より図を引用

その他手法 - 自己教師あり学習 –

- 異常画像が希少なため、大量な正常データから特徴を学習する教師なし学習がメインだった
- そこで、細かい傷や歪みのある異常画像を自ら作成
  - 小さい線を入れて傷を再現
  - 一部をくり抜いてずらして貼り付けることで歪みを再現
- 教師なしから教師ありへと設定が変わり、精度向上



(c) Cutout (d) Scar (e) CutPaste (f) CutPaste (Scar)

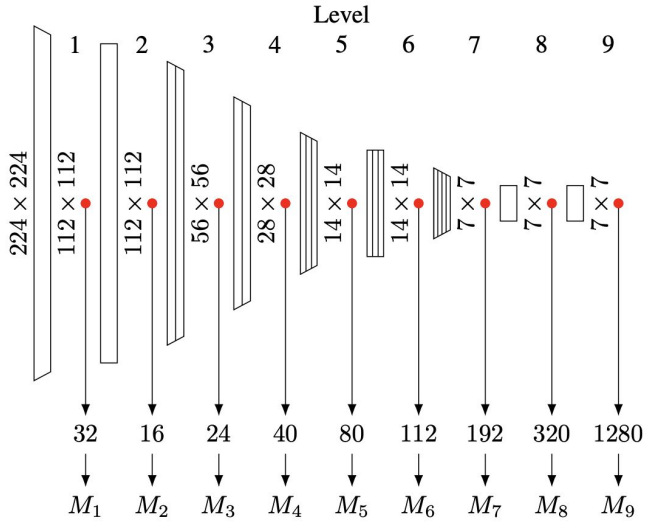
[23] Li, Chun-Liang, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. “CutPaste: Self-Supervised Learning for Anomaly Detection and Localization.” ArXiv:2104.04015 [Cs], April 8, 2021.より図を引用

その他手法 - 転移学習 –

- ImageNetのような大規模データセットで
  学習した分類モデル（Efficient Netなど）を
  異常検知に転用　→ **学習不要で早い！**

- 中間層の特徴量を用いて異常スコアを算出
  - 正常データの中間層特徴量が多変量正規分布に従っていると仮定
  - 分布の平均 μ と共分散行列 Σ から中間層特徴量xのマハラノビス距離を測り異常スコアとして利用

$$M(\mathbf{x}) = \sqrt{(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$



[24] ippel, Oliver, Patrick Mertens, and Dorit Merhof. “Modeling the Distribution of Normal Data in Pre-Trained Deep Features for Anomaly Detection.” ArXiv:2005.14140 [Cs], October 23, 2020.より図を引用

モデルの評価

評価指標

基本的な指標のおさらい

TP：実測と予測が両方陽性
FP：実際は陰性なのに、誤って陽性と判定
FN：実際は陽性なのに、誤って陰性と判定
TN：実測と予測が両方陰性

	True（実測）	False（実測）
True（予測）	TP	FP
False（予測）	FN	TN

Accuracy

- 全データに対する、正しい予測の割合

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

評価指標

不均衡なデータにはAccuracyが向いていない

誤って全てのデータを正常と判定した場合、次のような問題が発生する

- 均衡データ：正常 500件、異常 500件

  Accuracy = (0 + 500) / (0 + 0 + 500 + 500) = 0.5

- 不均衡データ：正常 990件、異常 10件

  Accuracy = (0 + 990) / (0 + 0 + 10 + 990) = 0.99

精度が良く見えてしまう！

均衡データ	異常（実測）	正常（実測）
異常（予測）	TP = 0	FP = 0
正常（予測）	FN = 500	TN = 500

不均衡データ	異常（実測）	正常（実測）
異常（予測）	TP = 0	FP = 0
正常（予測）	FN = 10	TN = 990

異常検知で主に使われる指標

Precision

- 陽性と予測したデータのうち、真の陽性の割合
- 誤検知の少なさを測定

$$Precision = \frac{TP}{TP + FP}$$

Recall（またはTrue Positive Rate）

- 全陽性データのうち、陽性と判定できた割合
- 見逃しの少なさを測る

$$Recall = \frac{TP}{TP + FN}$$

	True（実測）	False（実測）
True（予測）	TP	FP
False（予測）	FN	TN

Precision

Recall

見逃しと誤検知には異なるコストがかかり、どちらに重きを置くかは応用例によって異なる。

- 例）医療画像診断：見逃すと患者の生命に大きくかかわる一方で、誤検知は人力でカバーできるため、Recallを重視

異常検知で主に使われる指標

Precision・Recallをバランスよく見る指標

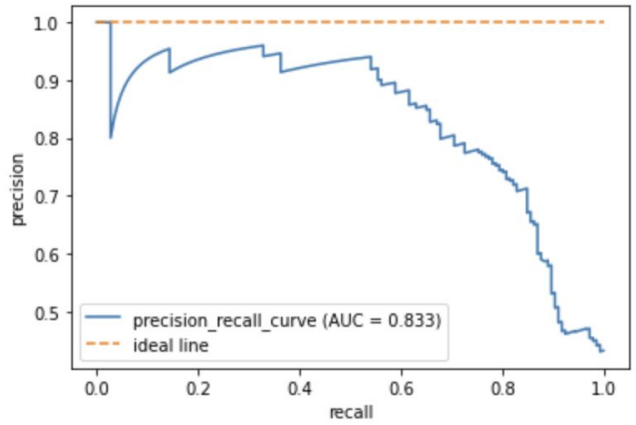
F値 (F1 score)

- PrecisionとRecallの調和平均
- バランスが良いほど大きな値に

$$F1\ score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

PR-AUC・ROC-AUC (Area Under the Curve)

- 曲線の下側の面積（最大1）で評価
- 面積が大きいほど良い

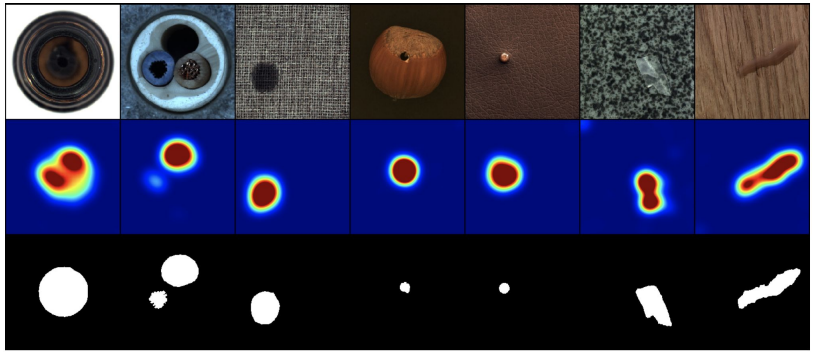


今後の課題と発展

今後の課題と発展

異常の解釈性

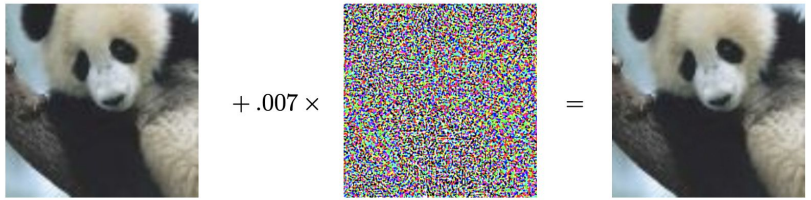
- 予測に基づいた意思決定の場面において、異常と判断した根拠は重要
- 異常・正常の分類と同時に、異常と判断した根拠を可視化するモデルが必要



[25] Liznerski, Philipp, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, Marius Kloft, and Klaus-Robert Müller. “Explainable Deep One-Class Classification.” ArXiv:2007.01760 [Cs, Stat], March 18, 2021. より図を引用

敵対的サンプル（Adversarial Example）

- モデルの誤分類を狙った意図的な異常は検知が難しい



[26] Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. “Explaining and Harnessing Adversarial Examples.” ArXiv:1412.6572 [Cs, Stat], March 20, 2015. より図を引用

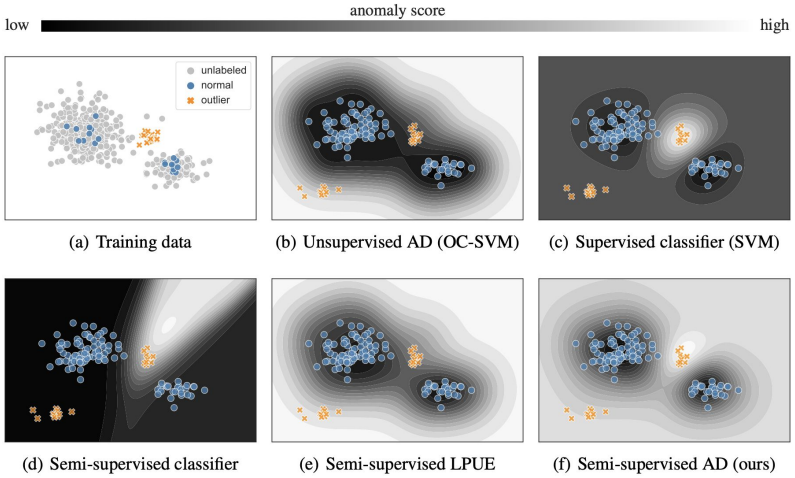
今後の課題と発展

少量のラベル付きデータの活用

- ラベル付き異常データが少量ながら手に入る状況も考えられる
- 少量のラベルを活用した半教師あり学習により教師なし学習より精度向上が見込める

ベンチマークデータセットの作成

- 既存の異常検知データセットは設定が簡単
  - 大きさ・向きが揃っているなど
- 発展にはさらに難易度の高いオープンデータセットが必要



[27] Ruff, Lukas, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. “Deep Semi-Supervised Anomaly Detection.” ArXiv:1906.02694 [Cs, Stat], February 14, 2020. より図を引用