

深層生成モデルによる異常検知

福水 健次 数理・推論研究系 教授

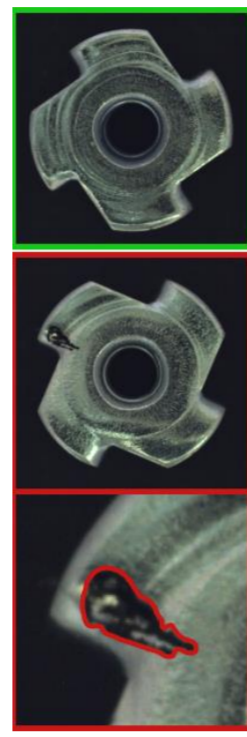
総研大・統計科学専攻・村瀬博典さんとの研究

参考文献: Murase, Fukumizu (2022). ALGAN: Anomaly Detection by Generating Pseudo Anomalous Data via Latent Variables. *IEEE Access*



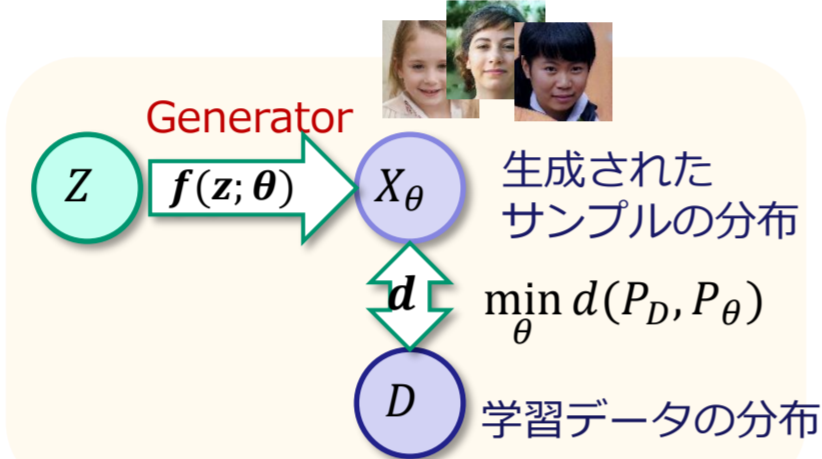
教師無し異常検知

- 異常検知
 - データから、正常/異常を識別する問題
 - 異常データを大量に集めるのは難しい (異常の発生は稀)
 - 「教師無し」異常検知: **正常データのみ**からの学習.
- 従来手法
 - 統計的手法: Z-score, 尤度, etc
 - 分布の疎密: kNN, local outlier factor (lof) など多くの手法
- 深層学習による異常検知
 - 画像の特徴抽出器として (ImageNet事前学習)
 - **深層生成モデル** (GANなど) による分布の学習 → 検知



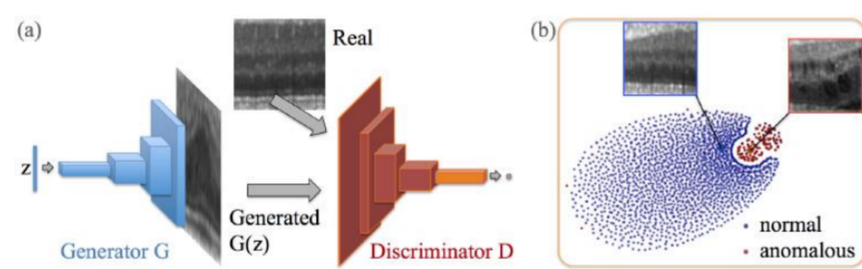
敵対的生成モデル (Generative Adversarial Nets, GAN)

- 訓練データの分布を近似するサンプルを生成するモデルを学習.
- 生成器: 正規分布のサンプルをNNにより変換 $\hat{X} = f_{\theta}(Z), Z \sim N(0, I_d)$
- 判別器: 生成データと訓練データの分布間距離を最大化 $\max E_{Data}[\log g_{\eta}(D)] + E_Z[1 - \log g_{\eta}(\hat{X})]$
- 敵対的学習 $\min_{\theta} \max_{\eta} E_{Data}[\log g_{\eta}(D)] + E_Z[1 - \log g_{\eta}(f_{\theta}(Z))]$



GANによる異常検知: 既存手法

- 直接的なGANの異常検知への応用
- 正常データを生成するGANを学習後、判別機 $D_w(x)$ で異常検出
- 問題点: 判別機は真のデータ/生成データという「近い分布」を判別するよう学習
- 異常な x に対する判別機の出力は不明
- GANによる異常検知の従来法: 潜在変数 Z に戻して (逆写像を構成) 比較する
- AnoGAN (Schlegl et al 2017), EGBAD (Zenati et al 2018), GANomaly (Ackay et al 2018)



提案手法: ALGAN

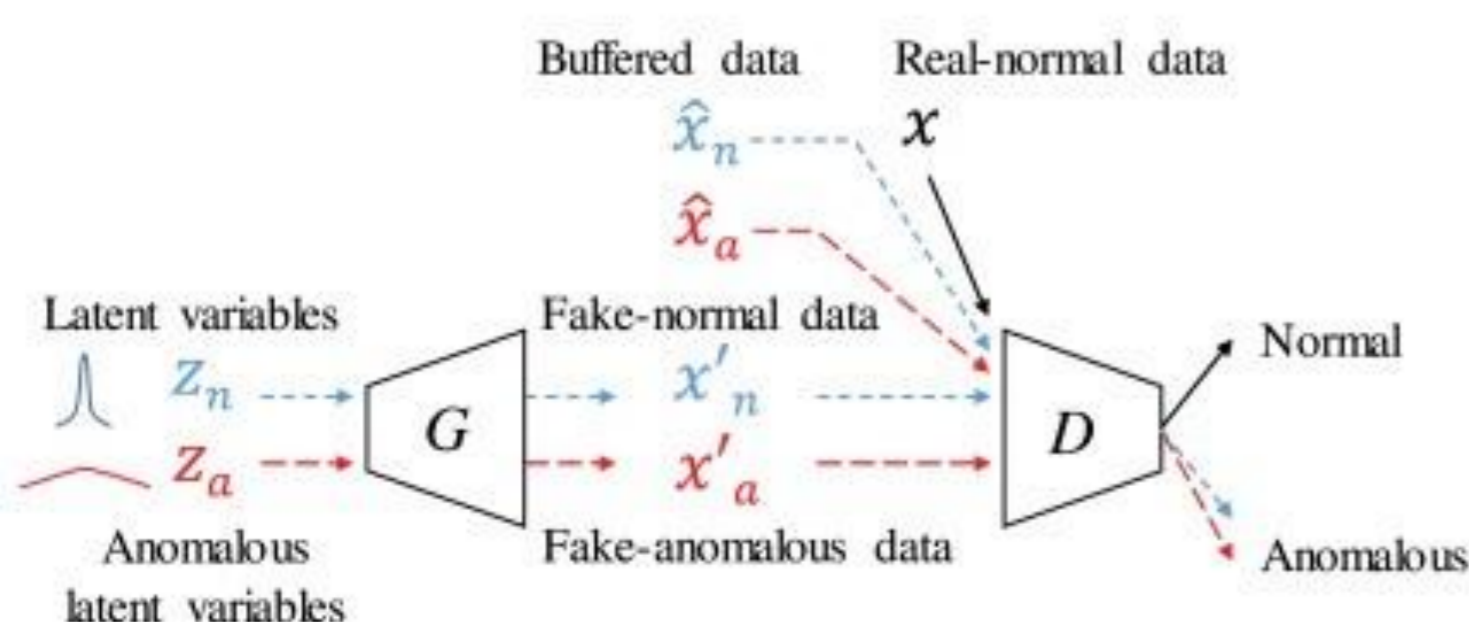
Anomalous Latent Variable GAN (ALGAN)

- 疑似異常データの生成
- 生成器の潜在変数 Z に異常潜在状態を加える
- $Z \sim \alpha N(0, I_m) + (1 - \alpha)N(0, \sigma^2 I_m)$ ($\sigma \gg 1$)
- 正常: $\mu - \delta_{\sigma}$ to $\mu + \delta_{\sigma}$
- 異常: $\mu - \delta_{\sigma}$ to $\mu + \delta_{\sigma}$ (with σ^2)

潜在変数 Z

生成器: 生成正常データ vs 訓練データ (正常) で学習
判別器: 生成データ (正常・異常) vs 訓練データを識別

学習が進む前の生成正常データも疑似異常に加える



ALGANの学習スキーム

Take-home Message:

ALGAN: 正常データのみからの異常検知法

- 深層生成モデルを使った疑似異常データ生成
- 任意のデータタイプに適用可能 (画像に限らない)
- 速い検知時間でSOTAに匹敵する検知性能

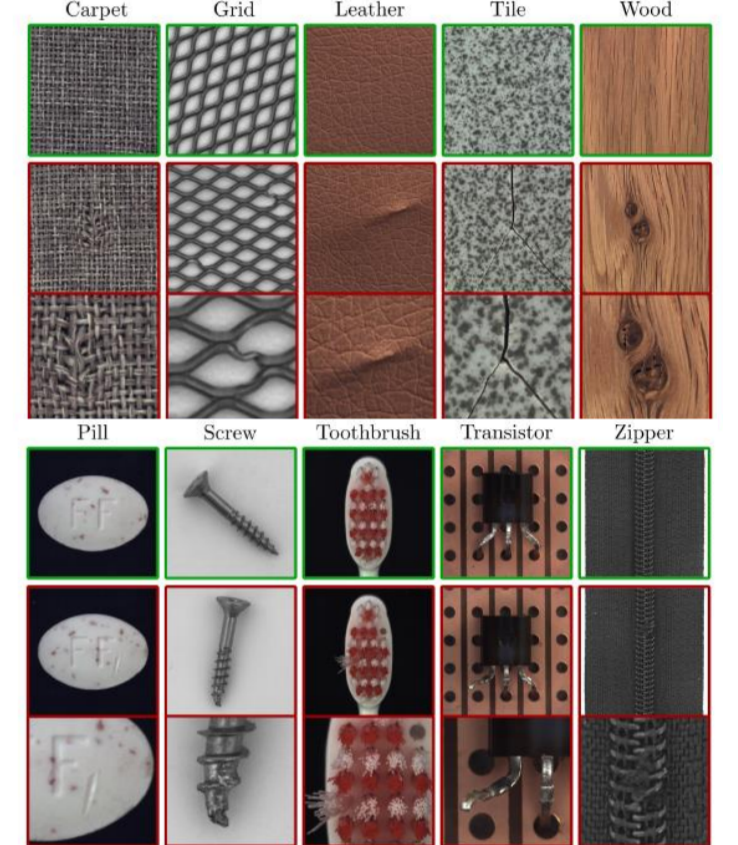
● 入力データは任意 c.f. 近年のSOTAは画像データに特化

- 画像データ または
- ImageNet事前学習による特徴

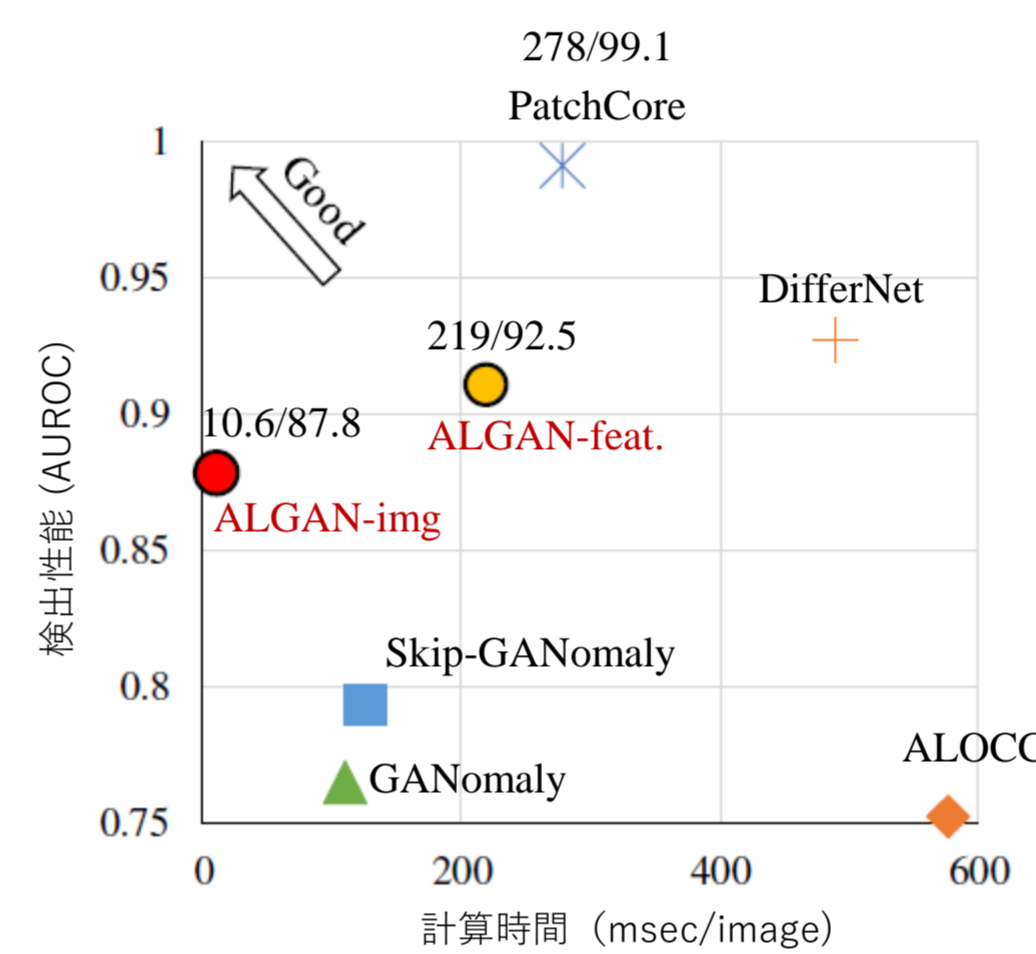
比較実験

MV-Tec AD (Bergmann et al. CVPR 2019)

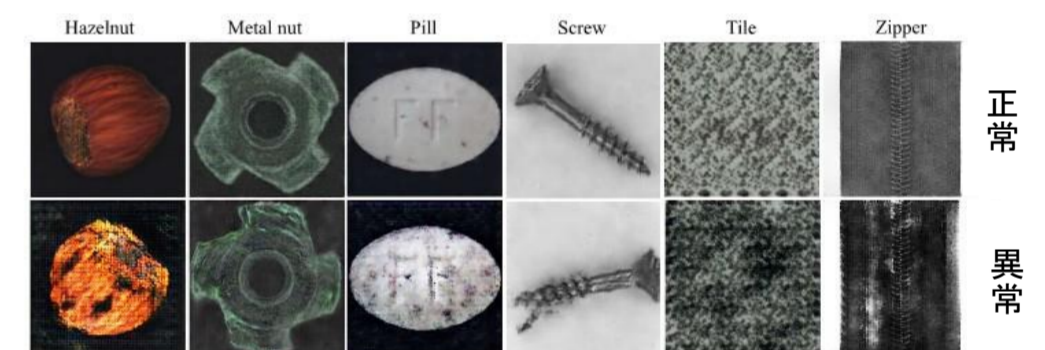
- 製品検査的な異常検知のための標準画像データセット
- Texture 5種/Object 10種, Total 5354画像
- 2つのデータタイプを利用
- a) ピクセルデータ 256 x 256
- b) WideResNet101による特徴



MVTec-ADの画像データ例



検知時間と検出性能 (AUROC)



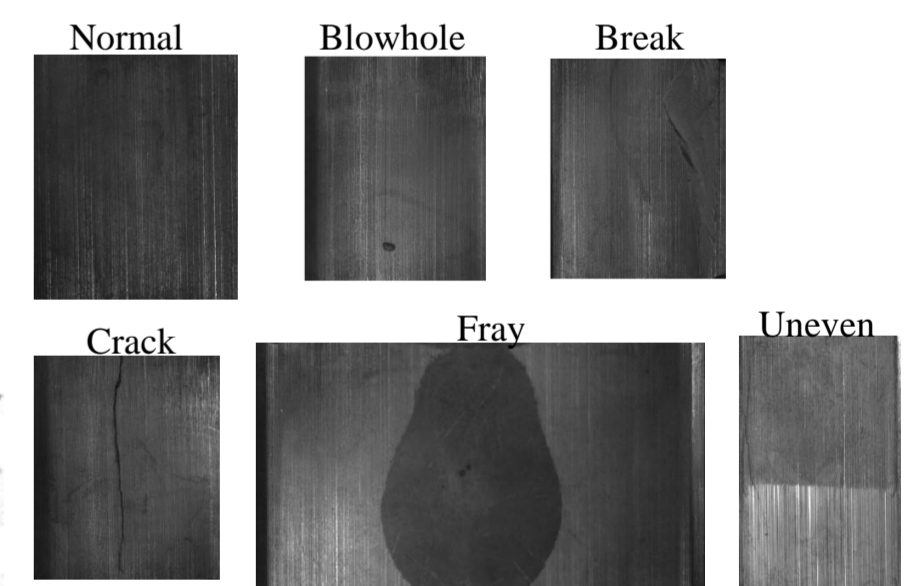
ALGANによる生成画像例

Magnetic Tile Defects (MTD)

- Data: 952正常/392異常

Training with image data			
AUROC	GANomaly	Skip-GANomaly	ALGAN-image
	0.683	0.504	0.956
Training with pre-trained features			
PatchCore-10	GANomaly-feat.	DifferNet	ALGAN-feature
0.979	0.766	0.977	0.824

AUROCの比較

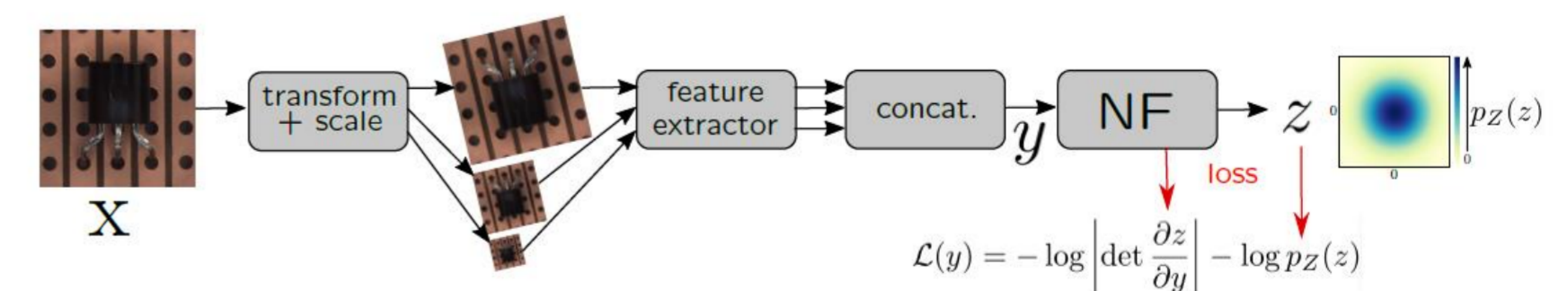


MTDデータの例

関連手法

DifferNet (Rudolf et al. WACV 2021)

- Normalizing Flow を使う → 尤度による異常判定
- AlexNetによる特徴と画素データを結合して用いる
- Data Augmentation: 回転, スケール変換.
- 計算量は大きい.



PatchCore (Roth et al. 2021)

- ImageNet+ResNetによる特徴を使う.
- 画素ごとの特徴を保存 → CoreSetを選択 (Greedy Search)
- kNNによって異常判定

