

Unsupervised Anomaly Detection based Generative Adversarial Networks (GAN)

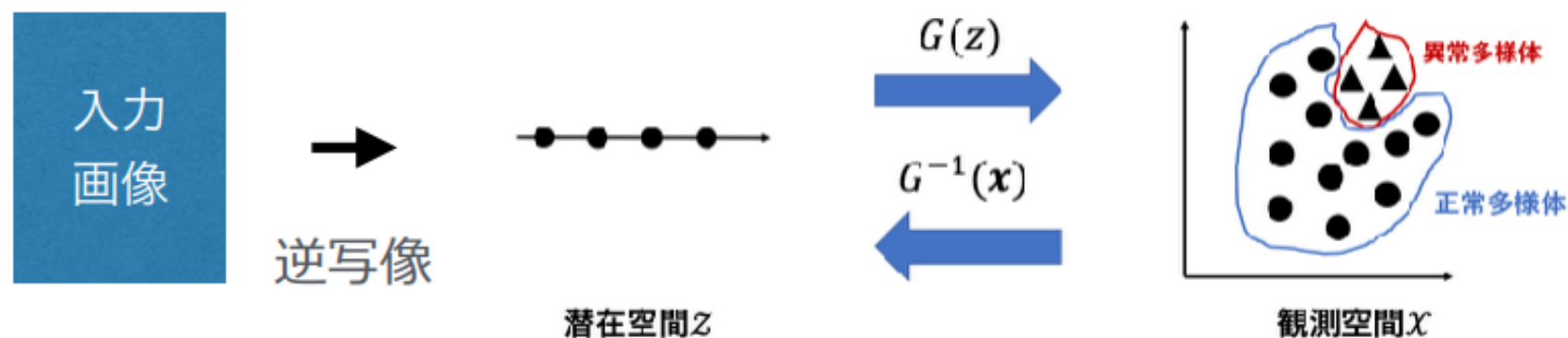
Le Van Cong

背景

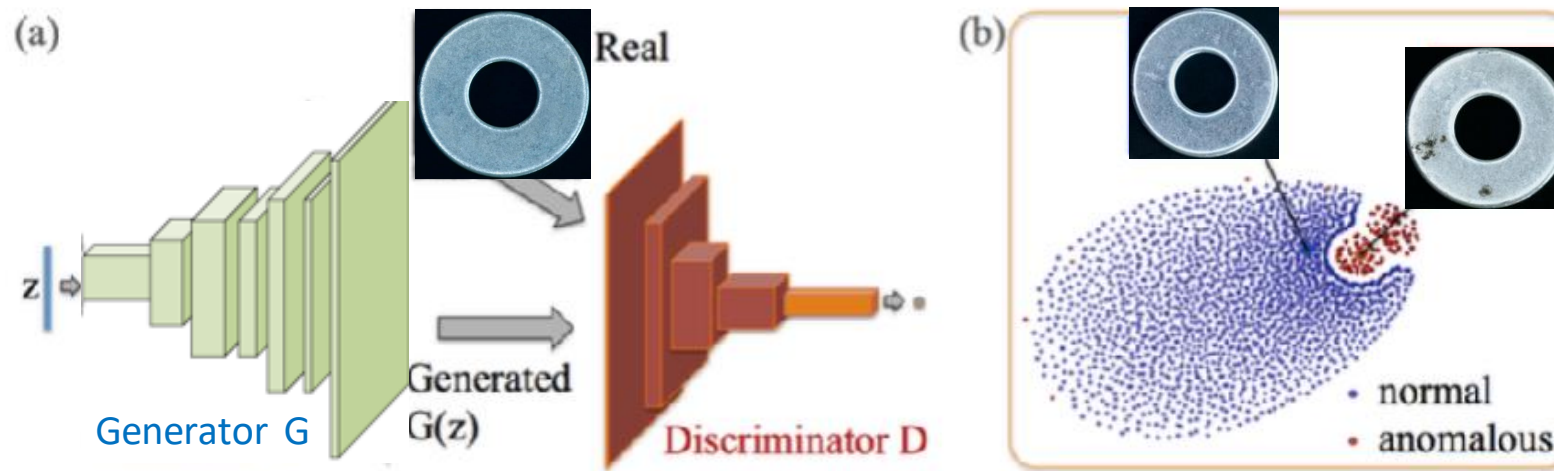
- CNN（Convolutional Neural Network）などの識別モデルでは、異常検知タスクにおいて異常データが少ないと
 - ・汎化性能が低い
 - ・学習・検証ができないなどの問題がある.
- GAN（Generative Adversarial Networks）などの生成モデルを用いた教師なし学習によって異常検知タスクを解決する

識別モデルと生成モデル

- 各クラスが分類可能であるとするならば，データの分布は異なる
- 識別モデル
 - ・ クラスとクラスを識別する境界線を探索
- 生成モデル
 - ・ 各クラスのどのように分布しているかを探索



学習アルゴリズム



- GANの価値関数

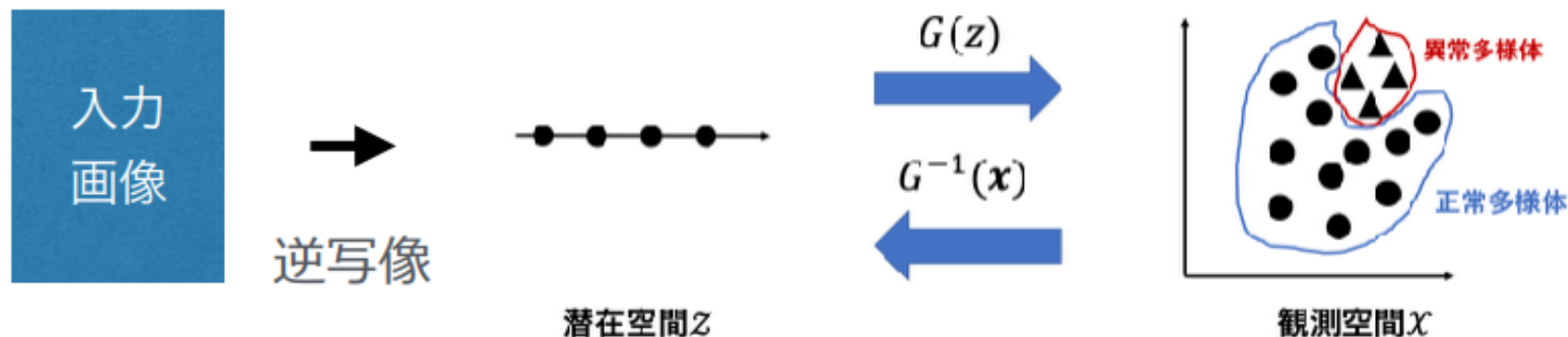
$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

- 通常のGAN同様，生成器は実際の画像に近いものを生成し，識別器は実際の画像と生成された画像を正しく識別する敵対学習を行う

学習アルゴリズム

- GANは潜在空間 z を観測空間に写像することはできるが、その逆写像は困難潜在空間 z は、以下の損失関数を最小化するを勾配法で探索する

$$\mathcal{L}(\mathbf{z}_\gamma) = (1 - \lambda) \cdot \mathcal{L}_R(\mathbf{z}_\gamma) + \lambda \cdot \mathcal{L}_D(\mathbf{z}_\gamma)$$



学習アルゴリズム

- 誤差関数
 - 残差損失関数

$$\mathcal{L}_R(\mathbf{z}_\gamma) = \sum |\mathbf{x} - G(\mathbf{z}_\gamma)|$$

- 識別損失関数

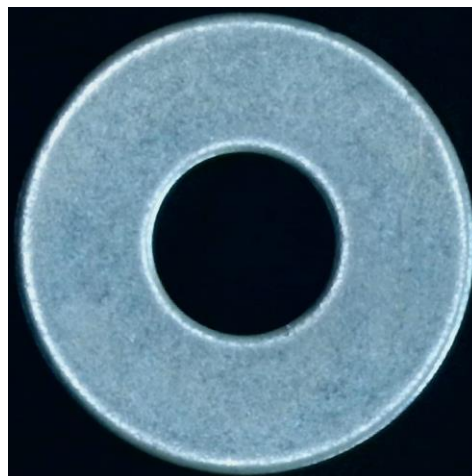
$$\mathcal{L}_D(\mathbf{z}_\gamma) = \sum |\mathbf{f}(\mathbf{x}) - \mathbf{f}(G(\mathbf{z}_\gamma))|$$

- 異常スコア

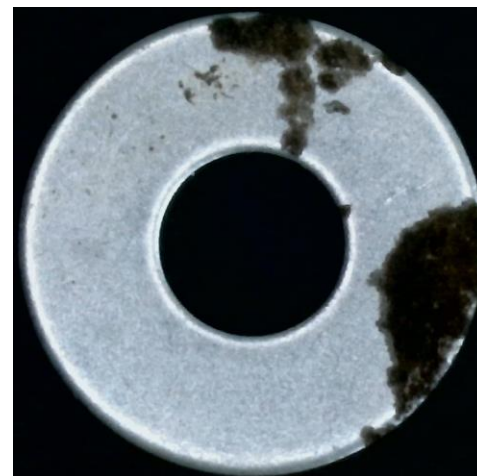
$$A(\mathbf{x}) = (1 - \lambda) \cdot R(\mathbf{x}) + \lambda \cdot D(\mathbf{x})$$

実験

- データセット
 - 網膜の正常, 異常データ
 - 64x64pxサイズ
 - 30 image ok -training
- パラメータ
 - 最適化手法: Adam
 - z-dim = 20
 - 5000 epochs



正常



異常

実験

- 定性的評価
 - 生成モデルがリアルな画像を生成できているか
 - 健康, 異常パターンを目視で確認
- 定量的評価
 - 異常検出精度の算出 残差スコア
 - 異常度・異常箇所ヒートマップ

実験結果

入力画像

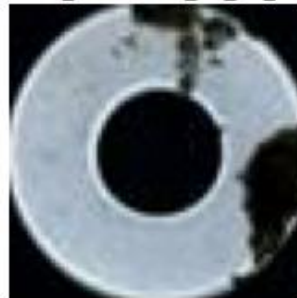
sabi-WIN_20200505_13_23_00_Pro



sabi-WIN_20200505_11_14_48_Pro



sabi-WIN_20200505_13_22_47_Pro



sabi-WIN_20200505_13_22_52_Pr



生成画像

recon-image -loss: 988



recon-image -loss: 885



recon-image -loss: 1002

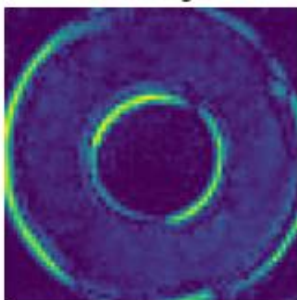


recon-image -loss: 1307

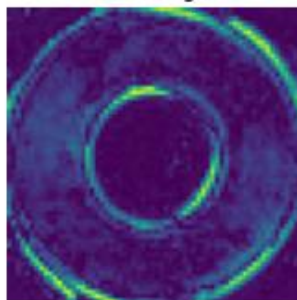


異常箇所の
ヒートマップ

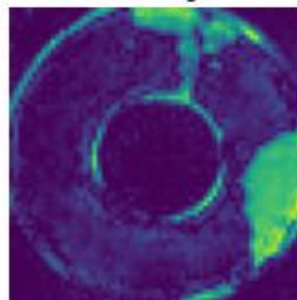
diff-image



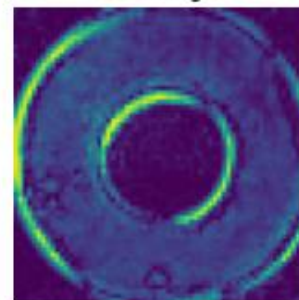
diff-image



diff-image



diff-image



結論

- リアルな正常画像を生成できている
- 画像に大きな異常を検出することができます
- 画像内の大きな異常と小さな異常の両方を検出できるように、さらにモデルを調査する必要があります
 - VAE
 - SkipGAN
 - GAN..