Ridge-i

ACCV 2018

Anomaly Detection Using GANs for Visual Inspection in Noisy Training Data

Masanari Kimura, Takashi Yanagihara

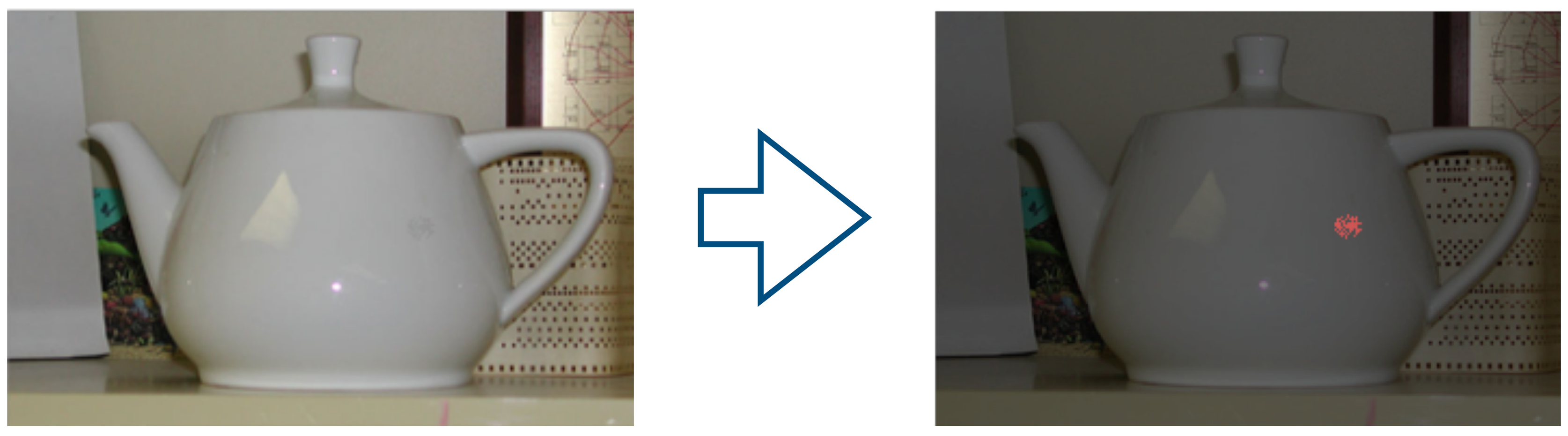
Motivation

Deep Learning based Anomaly Detection in Image

- VAE, GANs, etc.
- All of the training data are clean normal images only.

Anomaly Detection in Noisy Environments

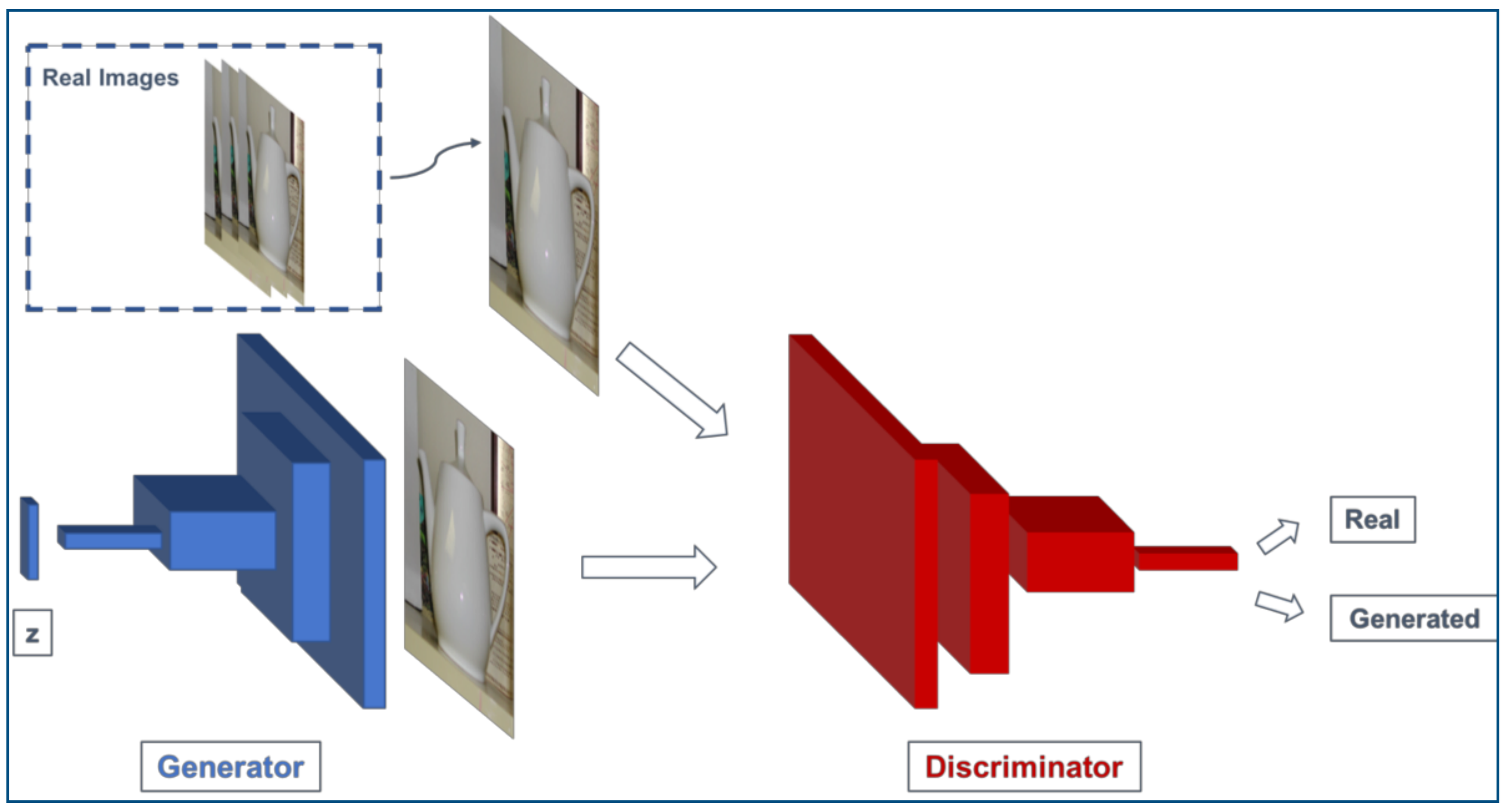
- In the real world, there are many cases where only noisy data can be used.
- We tackle the problem that the training data includes the "not" clean image.



Anomaly Detection Using GANs

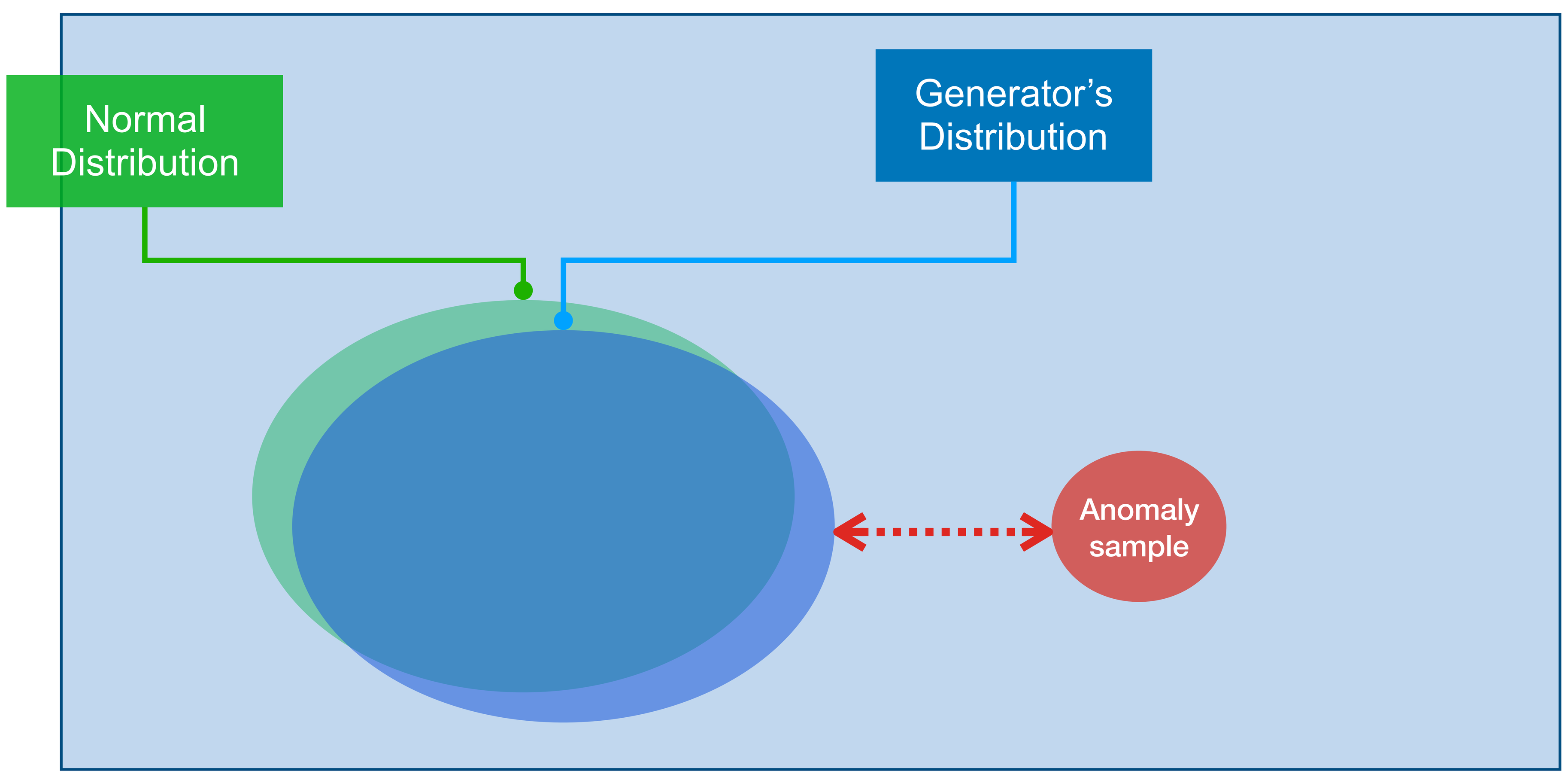
Training Generative Adversarial Network with Normal Data

- A generator learns to be able to make normal images.
- Discriminator learns to be able to distinguish real images and fake images.



Anomaly Detection Using Trained Generator

- Generate a fake sample most similar to the input sample and compare.
  - The generator can generate normal samples.
  - The generator can't generate anomaly samples.
- IIn the inference, we solve the optimization problem to search latent representation z.
  - Only the coefficients of z are adapted via backpropagation.
  - The trained parameters of the generator and discriminator are kept fixed.



Training Distribution Distortion

Method	Real samples	Fake samples
Original AnoGAN	Real images from dataset	Generated images
Noisy AnoGAN	Real images from dataset	Generated images & Anomaly samples

Our Objective

$$V'(D,G) = \gamma l_{Adv}(D,G) + (1-\gamma) l_{An}(D),$$

where

$$l_{Adv}(D,G) = V(D,G)$$
$$l_{An}(D) = \mathbb{E}_{x\sim p_{an}(x)}[\log(1-D(x))].$$

Theoretical Analysis

Original GANs objective function is following equation:

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

The equation can be transformed as follows:

$$V(D,G) = \int_x p_d(x)\log D(x) + p_g(x)\log(1-D(x))dx \quad (2)$$
$$V(D^*,G) = 2JSD(p_d||p_g) - 2\log 2 \quad (3)$$

Here, D^* is the optimum discriminator.
Our objective is,

$$V'(D,G) = \gamma l_{Adv}(D,G) + (1-\gamma) l_{An}(D), \quad (4)$$

where

$$l_{Adv}(D,G) = V(D,G) \quad (5)$$
$$l_{An}(D) = \mathbb{E}_{x\sim p_{an}(x)}[\log(1-D(x))] \quad (6)$$

We obtain the objective function of the generator as :

$$V'(D^*,G) = 2JSD(p_d||p_N) - 2\log 2 \quad (7)$$
$$p_N = \gamma p_g + (1-\gamma)p_{an}. \quad (8)$$

Since the JSD becomes the minimum when $p_d = p_N$, the optimum p_g^* for G can be derived as follows:

$$p_N^* = p_d \quad (9)$$
$$\gamma p_g^* + (1-\gamma)p_{an} = p_d \quad (10)$$
$$\gamma p_g^* = p_d - (1-\gamma)p_{an} \quad (11)$$
$$p_g^* = \frac{1}{\gamma}p_d - \frac{(1-\gamma)}{\gamma}p_{an} \quad (12)$$

where $\frac{1}{\gamma}\Gamma \geq 1$ and $\frac{(1-\gamma)}{\gamma} = \frac{1}{\gamma} - 1 \geq 1$.

we may consider that the proposed method distorts the distribution of real images to remove anomaly images and to make the ideal distribution of normal images.

Experimental Results

Experimental results for IR-MNIST

- The training dataset doesn't contain class "3".
- The top images are the original test images.
- In the test image, "3" exists unlike the training image.

Original samples

Differences