

<https://doi.org/10.7236/JIIBC.2020.20.3.171>  
JIIBC 2020-3-24

## 이미지의 질과 왜곡을 고려한 적대적 생성 신경망과 이를 이용한 비정상 검출

### Anomaly Detection of Generative Adversarial Networks considering Quality and Distortion of Images

서태문\*, 강민국\*\*, 강동중\*\*\*

Tae-Moon Seo\*, Min-Guk Kang\*\*, Dong-Joong Kang\*\*\*

**요 약** 최근 연구 결과에 따르면, 컨볼루션 신경 회로망은 이미지 분류, 객체 검출, 이미지 생성 등의 문제에서 최고의 성능을 보여주고 있다. 비전 카메라를 사용한 결함 검사는 다른 결함 검사보다 경제적이기 때문에 공장 자동화에 있어서 아주 중요하고, 딥러닝의 지도학습은 전통 기계학습 방식의 결함 검사 성능을 월등히 뛰어넘었다. 하지만, 딥러닝의 지도 학습은 엄청난 양의 데이터 주석 작업을 요구하기 때문에, 이를 실제 산업 현장에 적용하는 것은 효율적이지 않다. 따라서 본 연구는 최근 이미지 생성 과업에서 큰 성공을 보여주고 있는 변분 오토인코더와 적대적 생성 신경망을 활용하여 비지도 방식의 비정상 검출을 위한 신경망 회로 구조를 제안하였고, 이를 MNIST, 용접 결함 데이터에 적용하여 비정상 검출 성능을 검증하였다.

**Abstract** Recently, studies have shown that convolution neural networks are achieving the best performance in image classification, object detection, and image generation. Vision based defect inspection which is more economical than other defect inspection, is a very important for a factory automation. Although supervised anomaly detection algorithm has far exceeded the performance of traditional machine learning based method, it is inefficient for real industrial field due to its tedious annotation work, In this paper, we propose ADGAN, a unsupervised anomaly detection architecture using the variational autoencoder and the generative adversarial network which give great results in image generation task, and demonstrate whether the proposed network architecture identifies anomalous images well on MNIST benchmark dataset as well as our own welding defect dataset

**Key Words** : Anomaly Detection, Unsupervised learning, Generative Adversarial Network, Variational Autoencoder

\*정회원, 부산대학교 기계공학과

\*\*정회원, 포항공과대학교 컴퓨터공학과

\*\*\*정회원, 부산대학교 기계공학과

접수일자 2020년 4월 21일, 수정완료 2020년 5월 21일  
게재확정일자 2020년 6월 5일

Received: 21 April, 2020 / Revised: 21 May, 2020 /

Accepted: 5 June, 2020

\*Corresponding Author: dj kang@pusan.ac.kr

Dept. Mechanical Engineering, Pusan University, Korea

## 1. 서 론

### 1. 개요

비정상 검출은 다수의 내좌측(Inlier) 데이터와 이상점(Outlier) 데이터가 존재하는 상황에서 이상점 데이터를 검출하는 과정이다. 내좌측 데이터란 풍부한 양의 정상적인 데이터를 의미하고, 이상점 데이터란 제품 결함 사진과 같이 내좌측 데이터에 비해 그 수가 매우 적고 부적절한 데이터를 의미한다. 전통적으로 비정상 검출은 확률 분포를 근사하는 통계적 모형화<sup>1,2</sup>를 통해서 수행되어 왔는데, 최근 다양한 경로에서 엄청난 양의 데이터 들이 수집되고, 그 형태가 점점 더 복잡해 짐에 따라 통계적 모형을 통한 비정상 검출은 차원의 저주에 빠져 제 역할을 하지 못하게 되었다. 하지만, 해결이 불가능해보였던 차원의 저주를 극복하고 이미지 분류<sup>3</sup>, 객체 검출<sup>4</sup>, 이미지 생성<sup>5</sup> 등에서 최고의 성능을 내고 있는 딥러닝의 등장으로 인해 다양한 딥러닝 기반 비정상 검출 알고리즘 들이 제안<sup>6,7,8,9</sup>되었고 이들은 다른 분야와 마찬가지로 전통적인 비정상 검출 알고리즘의 성능을 제치고 최고의 성능을 보여주고 있다.<sup>9</sup>

최신 딥러닝 기반의 비정상 검출은 오토인코더와 신경망 규제 방법을 통해 데이터에 대한 풍부한 정보를 가지고 있는 잠재변수를 추출한 후, 이를 SVM(Support Vector Machine) 분류기에 인가하여 비정상을 검출하는 방법<sup>9</sup>, 변분 오토인코더와 같은 확률적 오토인코더를 이용하는 방법<sup>10</sup> 적대적 생성 신경망<sup>11,12</sup>을 통해 정상 이미지에 대한 확률분포를 학습하고, 학습에 참여하지 못한 비정상 데이터가 신경망에 인가되면, 왜곡된 이미지가 생성되는 현상을 이용한 방법이 있다.

본 논문에서는 지각(Perceptual)변분 오토인코더<sup>13</sup>와 경계 평형 적대적 생성 신경망<sup>14</sup>의 영감을 받아 이미지의 품질과 왜곡을 모두 고려하여 비정상 여부를 결정하는 ADGAN을 제안한다. ADGAN은 총 3개의 부분으로 나뉘는데, 순서대로 개선 및 악화부, 생성부, 검출부이다(See Figure 1). 지각 변분 오토인코더로 구성되어 있는 개선 및 악화부는 다수의 정상 이미지에 대한 잠재변수를 확률적으로 추출하여 기존의 오토인코더를 사용한 추출 방식보다 더욱 표현력이 좋은 다양체 공간(Manifold space)을 구축하게 된다. 따라서 개선 및 악화부는 네트워크에 인가된 정상 이미지를 생성부가 학습할 이미지 공간으로 사상(mapping)시키고, 비정상 이미지의 경우는 반대의 역할을 수행하여 생성부에 인가될 이미지를 개선 및 악화시키는 역할을 수행한다. 다음으로 생성부와

검출부는 서로를 속이기 위한 적대적 학습을 하는데, 검출부의 경우, 실제 이미지(정상)와 생성된 이미지를 입력으로 받아, 해당 이미지가 원본 인지 생성된 이미지인지를 판단할 수 있도록 학습되고, 생성부는 개선 및 악화부에서 생성된 이미지를 입력으로 받아 검출부가 실제 이미지(정상)인 것으로 착각할만한 이미지를 생성하도록 학습되어 진다. 따라서, 제안된 구조에 정상 이미지가 인가되면 검출부가 헛갈릴만한 이미지를 생성하게 되지만, 비정상적인 이미지가 인가된다면 개선 및 악화부에서 왜곡된 이미지를 생성할 것이고, 결과적으로 생성부가 생성한 이미지가 검출부를 속이지 못하여 비정상 판정을 받게 된다.

요약하자면, 이 논문의 기여점은 다음과 같다. (1) 지각 변분 오토인코더<sup>13</sup>와 적대적 생성 신경망<sup>14</sup>을 사용하여 비정상 검출 과업에서 최고 성능을 내는 구조를 제안하였다. (2) 기존 비정상 감지에서 사용했던 픽셀 단위의 L1, L2 distance 목적함수에 최근 많은 주목을 받고 있는 Wasserstein distance를 추가하여, 입력 이미지와 출력 이미지의 픽셀 단위 거리뿐만 아니라, 이미지의 질적인 면까지 고려하는 목적함수를 제안하였다. (3) 앞에서 제안한 목적함수를 이용하여 비정상 요소가 존재하는 부분을 가시화하는 방법을 제안하였고, 용접 결함 데이터를 통해 실험적으로 검증하였다.

### 2. 기존 연구사례

#### 가. Statistical Modeling

신경망 회로를 이용한 비정상 검출 알고리즘이 등장하기 이전에는 비정상적인 데이터를 검출하기 위해 통계적 모형화를 주로 사용하였다. 통계적 모형화에는 주어진 데이터의 표본평균( $\bar{X}$ )과 표본분산( $S^2$ )을 통해 계산된 검정 통계량이 제시된 신뢰구간에 속하지 않으면 비정상으로 판정하는 카이제곱 검정 및 티 검정이 있다. 또한, 공분산 행렬을 이용하여 주어진 데이터를 차원 감소시킨 후, 선형 분류기에 인가하여 비정상 여부를 검출하는 강건 주 성분분석(RPCA)<sup>17</sup>, 데이터를 특정 확률 분포에 적합시킨 후 데이터와 확률 분포 사이의 거리를 측도로 삼아 비정상을 판정하는 방법<sup>2</sup> 등이 제안되어 왔다.

#### 나. Deep Learning based Anomaly Detection

컨볼루션 신경망 회로(Convolutional Neural Networks, CNNs)가 데이터의 특징 공간을 풍부하게 학습할 수 있음이 실험적으로 증명됨에 따라<sup>12,13,18</sup> CNN 기반의 다양

한 비정상 검출 알고리즘들이 제안되어 왔다. 방법론은 크게 두 가지로, 첫 번째는 오토인코더의 풍부한 특징 공간 학습력을 이용한 방법<sup>6,9,10</sup>이고, 두 번째는 적대적 생성모델을 활용한 방법<sup>7,8,19</sup>이다. 오토인코더는 데이터의 잠재변수를 찾기 위해 데이터의 차원을 축소(Encoding)시킨 후, 처음의 차원으로 다시 복원(Decoding)시키는 방법론인데, 축소되었던 저차원의 특징을 데이터의 잠재변수 공간으로 가정하고, SVM 또는 선형 분류기에 인가하여 이상점 검출을 하는 방법이 오토인코더를 이용한 비정상 검출 방법이다. 하지만, 최근 적대적 생성 신경망이 이미지 생성 및 특징 추출 과업에서 성공적인 성능을 보여줌에 따라<sup>12,14,20,21,22,23</sup> 적대적 생성 신경망의 학습을 통한 비정상 검출 알고리즘들이 많이 제안되고 있다. 이는 정상 이미지에 대한 확률 분포를 학습한 후, 정상 이미지 분포에서 거리가 먼 비정상 이미지가 인가되었을 때, 왜곡된 이미지가 생성되는 현상을 이용한 것으로, 인가된 이미지와 출력된 이미지의 L1 또는 L2 거리를 비교해 비정상을 검출한다.

### 3. 연구목표 및 범위

본 연구는 정상 이미지만을 사용하여, 학습과정에 등장하지 않는 비정상 이미지를 검출하기 위한 알고리즘을 제안하고, 이에 대해 해석적, 실험적 분석을 하는 것을 본문의 범위로 하였다.

## II. 이론적 배경

본 챕터에서는 논문에서 제안한 3가지 구조인 개선 및 약화부, 생성부, 검출부의 역할과 목적함수, 최적화 방법, 비정상 점수 계산법 및 비정상 가시화 방법에 대해서 설명한다.

### 1. 개선 및 약화부

많은 양의 데이터를 다룰 때 데이터의 특징을 파악하는 것은 아주 중요한 일이다. 딥러닝은 다양한 방식의 데이터 특징추출 방법을 제안해왔는데, 그 중 오토인코더<sup>24</sup>가 많은 분야에서 성공적으로 사용되고 있다. 또한, 오토인코더는 이미지 생성 과업에도 사용되는데 매 학습마다 추출되는 이미지의 특징 공간이 바뀌고, 입력 이미지를 단순히 항등 함수에 인가한 것과 같은 결과물을 생성하는 단점으로 인해 잘 사용되지 않고 있다. 오토인코더의 이러한 단점을 보완한 것이 지각 변분 오토인코더(PVAE)이다. 지각 변분 오토인코더의 시초인 변분 오토인코더(VAE)는 관측된 데이터를 가장 잘 설명하는 모형의 모수를 찾아주는 방법인 최대 우도 추정법(Maximum Likelihood Estimation, MLE)을 변분 추론(Variational Inference)과 최급 강하법(Gradient Descent)을 이용해 수치 해석적으로 해결한 것인데, 이때 입력 데이터와 출력 데이터 사이의 Cross-Entropy 목적함수를 미리 학습된 VGG19<sup>18</sup>의 지각(Perceptual) 목적 함수로 대체 한 것이 PVAE이고,

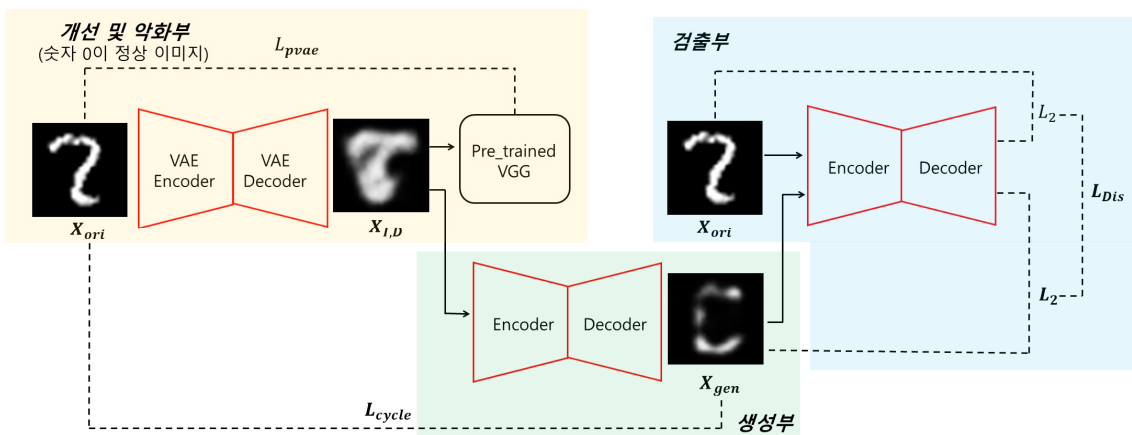


그림 1. ADGAN 개요

Fig. 1. ADGAN overview: When normal images are given to the Improver & Deteriorator, They try to generate data located in the input space of the Generator. However if abnormal images are given, I & G will try to reduce the quality of the generated images(Left Top). Generator and Discriminator(Middle bottom, Right Top) are adversarially trained to deceive each other and eventually Discriminator gets the ability to calculate the abnormal score from data.

그 목적 함수는 아래와 같다.

$$L_{PVAE} = D_{KL}(q_\theta(Z|X) \| P_\theta(Z)) + P(X_{ori}, X_{LD}) \quad (1)$$

$$\begin{aligned} \text{Where, } P(X, Y) = & MSE(VGG_1(X), VGG_1(Y)) \\ & + MSE(VGG_2(X), VGG_2(Y)) \\ & + MSE(VGG_3(X), VGG_3(Y)) \end{aligned}$$

$$VGG_n(X) = \text{미리 학습된 VGG19의 } n\text{번째 활성맵}$$

PVAE는 학습과정에 참여한 정상 이미지에 대한 잠재 공간을 밀도 있게 구축하여, 학습과정에 등장하지 않았던 새로운 정상 데이터의 잠재변수는 특징공간의 중심 사상해주고, 비정상 데이터의 경우는 공간의 중심에서 멀게 사상해줌으로서 정상 이미지를 입력으로 받았을 때는 정상적으로, 비정상 이미지를 입력으로 받았을 때는 왜곡된 이미지가 생성되도록 만들어주는 역할을 한다.

## 2. 생성부와 검출부

적대적 생성 신경망(Generative Adversarial Networks, GANs)은 생성된 이미지의 확률분포를 생성부와 검출부의 적대적인 학습을 통해, 생성하고 싶은 이미지의 확률 분포로 근사시켜 실제와 같은 이미지를 생성하는 방법으로 그 목적함수는 아래와 같다.

$$\begin{aligned} \min_G \max_D V(D, G) = & E_{x \sim p_{\text{data}}} [\log(D(X))] \\ & + E_{z \sim p_z} [\log(1 - D(G(z)))] \quad (2) \end{aligned}$$

$$\begin{aligned} \text{where, } G(X): & X\text{를 생성부에 인가한 결과 값} \\ D(X): & X\text{를 검출부에 인가한 결과 값} \\ & 0 \leq D(A) \leq 1 \end{aligned}$$

이때, 이미지의 위조 여부를 판별해주는 검출부의 출력 값인  $D(A)$ 가 1에 가까우면, A를 실제 이미지로,  $D(A)$ 가 0에 가까우면, A를 생성된 이미지로 판별한다. 이러한 적대적 생성 신경망은 이미지의 위조 여부를 검출할 때, 고차원 이미지에 대한 픽셀 단위의 평균 제곱 오차를 계산하는 것이 아닌, 생성된 이미지가 실제 이미지와 질적(Divergence 관점)으로 얼마나 다른지를 수치화하여 위조 여부를 결정하게 된다. 본 논문에서 제안하는 생성부와 검출부는 개선 및 악화부에서 생성된 이미지를 픽셀 단위로 비교할 뿐만 아니라, 질적인 비교를 모두 수행하여 효율적인 비정상 검출을 수행하게 된다.

## 3. 단방향 순환 목적함수

### (Unidirectional Cycle Consistent Objective)

적대적 생성 신경망의 엄청난 성공에도 불구하고, 적대적 생성 신경망을 응용한 기술이 실제 산업현장에서 사용되지 못하는 이유는 생성부가 생성하는 이미지가 검출부를 쉽게 속이는 특정 모드(Mode)에 갇혀 다양한 이미지를 생성 하지 못하게 되는 모드 붕괴(Mode Collapse)현상 때문이다. 모드 붕괴 현상을 해결하기 위해 다양한 종류의 신경망이 제안되어 왔지만<sup>12,14,20,21,22</sup> 여전히 근본적인 해결책을 찾지 못하고 있으며, 현재는 위에서 제안된 구조에 여러 가지 규제기법을 추가하여 모드 붕괴에 강건한 생성 구조를 구축하는 것을 해결책으로 삼고 있다. 따라서 본 논문에서는 모드 붕괴에 강건하고, 비정상 감지에 적합한 목적함수인 단방향 순환 목적함수를 제안하고, 그 수식은 아래와 같다.

$$L_{\text{cycle}} = \frac{MSE(X_{ori}, X_{gen})}{2} \quad (3)$$

### 베르스타인 목적함수를 이용한 적대적 생성 신경망의 목적함수

적대적 생성 신경망은 모드 붕괴 문제로 인해 수렴이 몹시 어려운 단점을 가지고 있다. 이는 생성부가 고정되어 있을 때, 적대적 생성 신경망의 수렴을 위해선 Jensen-Shannon divergence(JSD)의 값을 줄이는 방향으로 최적화를 해줘야 하기 때문인데<sup>21</sup> 이에 대한 해결책으로 JSD보다 수렴시키기 편한 베르스타인 거리(Wasserstein distance)를 목적함수로 사용한 적대적 생성 신경망이 많이 제안되고 있다.<sup>5,14,21,25</sup> 특히 2017년에 제안된 경계 평형 적대적 생성 신경망(BEGAN)<sup>14</sup>은 3채널의  $128 \times 128$  해상도 이미지를 생성할 수 있고, 경계 학습이라는 새로운 방법을 통해 학습이 매우 안정적이므로 본 논문에서 제안한 ADGAN의 기본 구조로 사용하였고, 그 목적함수는 아래와 같다.

$$G_{\text{loss}} = 2 \times MSE(X_{gen}, D(X_{gen})) \quad (4)$$

$$D_{\text{loss}} = 2 \times MSE(X_{ori}, D(X_{ori})) - k \times G_{\text{loss}} \quad (5)$$

$$\text{where, } \gamma = \frac{E[D(X_{gen})]}{E[D(X_{ori})]}$$

$$k_{t+1} = k_t + \lambda(\gamma D(X_{ori}) - D(X_{gen}))$$

### 제안된 구조의 최적화 방법

제안된 구조의 학습을 위해서는 총 3가지의 목적함수가 필요하고 수식을 나열하면 아래와 같다.

$$L_{PVAE} = D_{KL}(q_{\theta}(Z|X) \| P_{\theta}(Z)) + P(X_{ori}, X_{I,D}) \quad (6)$$

$$L_{ADGAN\_G} = L_{cycle} + G_{loss} \quad (7)$$

$$L_{ADGAN\_D} = D_{loss} \quad (8)$$

수식 (6)은 지각 변분오토인코더의 학습을 위한 목적함수이고, 수식 (7)과 수식 (8)은 생성부와 검출부의 적대적 학습을 위한 목적함수이다. 따라서 제안된 ADGAN을 최적화하기 위해서는 위의 수식 (6), (7), (8)을 적절한 최적화 알고리즘을 사용하여<sup>26,27</sup> 1:1:1 비율로 가중치 업데이트 해 줘야 한다.

### 비정상 점수 계산법

본 논문은 입력 이미지와 출력 이미지의 픽셀 단위 거리뿐만 아니라, 이미지의 질적인 면까지 고려하는 비정상 점수를 아래와 같이 제안한다.

$$S_{Anomaly} = \quad (9)$$

$$P(X_{ori}, X_{I,D}) + MSE(X_{I,D}, X_{gen}) + L_{cycle} + (2MSE(X_{gen}, D(X_{gen})) - 2MSE(X_{ori}, D(X_{ori})))^2$$

비정상 점수는 총 4가지 소 목적함수로 구성되는데, 계산 순서대로 설명하면,  $P(X_{ori}, X_{I,D})$ 는 개선 및 악화부에서 생성된 이미지가 입력 이미지와 지각적으로 (Perceptual) 얼마나 다른지를 나타내는 역할을 한다. 두 번째 목적함수인  $MSE(X_{I,D}, X_{gen}) + L_{cycle}$ 는 네트워크에 인가된 이미지, 개선 및 악화부를 통과하여 생성된 이미지, 생성부가 생성한 이미지의 픽셀 단위의 거리를 수치화한 것으로 이미지의 왜곡 정도를 나타내는 역할을 한다. 마지막으로

$[2MSE(X_{gen}, D(X_{gen})) - 2MSE(X_{ori}, D(X_{ori}))]^2$ 는 생성된 이미지와 원본 이미지의 Wasserstein 거리이고 이는 생성부에서 생성된 이미지와 원본 이미지의 확률 분포상 거리를 의미한다.

제안된 네트워크에 비정상 이미지가 입력되면, 개선 및 악화부는 입력 이미지의 잠재변수를 정상 이미지의 다양체 공간과 먼 곳으로 사상시킨 후, 왜곡된 이미지를

생성하여 높은  $P(X_{ori}, X_{I,D})$ 값을 가지게 한다. 이후 왜곡된 이미지가 생성부에 인가되면, 입력된 이미지가 정상적인 이미지의 확률 공간에서 추출된 것이 아니기 때문에, 더욱 왜곡된 이미지를 생성하게 된다. 따라서 생성부를 통해 생성된 이미지는 원본과 매우 다르므로(See Figure 1) 네트워크에 인가된 이미지, 개선 및 악화부를 통과하여 생성된 이미지, 생성부가 생성한 이미지 사이의 픽셀 단위 거리를 나타내는  $MSE(X_{I,D}, X_{gen}) + L_{cycle}$ 는 높은 값을 가지게 된다. 마지막으로,

$(2MSE(X_{gen}, D(X_{gen})) - 2MSE(X_{ori}, D(X_{ori})))^2$ 는 두 번 왜곡된 이미지와 원본 비정상 이미지 사이의 Wasserstein 거리를 나타낸다. 왜곡된 이미지와 원본 비정상 이미지 사이의 확률 분포상 거리가 멀기 때문에 높은 값을 가지게 된다.

### 비정상 가시화 방법

위에서 제안한 비정상 점수는 입력 이미지의 픽셀 별 비정상 점수를 계산한 후, 평균을 구해준 것과 같다. 따라서 평균 제곱 오차 연산을 제곱 오차 연산으로 바꿔주면 픽셀 별 비정상 점수를 계산할 수 있고, 이를 적용한 결과는 아래 실험의 Figure 5와 같다.

## III. 실험 및 결과

본 논문은 제안된 알고리즘의 성능을 확인하고, 개선 및 악화부, 생성부, 검출부의 효율성을 검증하기 위해 다음의 3가지 실험을 수행하였다. (1) 벤치마크 데이터 세트인 필기체 데이터를 통한 비정상 검출, (2) 필기체 데이터를 사용한, 개선 및 악화부, 생성부, 검출부에 적용된 목적함수의 유용성 검증, (3) ADGAN을 이용한 용접 결합 검사.

실험(1)의 결과 다음 장의 Figure 2에서 확인할 수 있듯, ADGAN이 State of the art인 DSVDD<sup>9</sup>의 성능을 7가지 실험 환경에서 능가했음을 확인하였고, 실험(2)의 결과인 Figure 3는 본 논문에서 제안한 여러 가지 목적함수가 비정상 감지에 잘 부합하는 목적함수임을 실험적으로 보여주고 있다. 마지막으로 ADGAN을 실제 산업 현장의 용접 결합 사진에 적용하였는데, 그 결과는 0.924 AUC이다. 이는 실제 상용화 가능한 수준(98% 이상)에는 못 미치지만, ADGAN을 이용한 비정상 검출이 어느 정도 상용화 가능 단계까지 진입했음을 실험적으로 보여주고 있다.

표 1. MNIST 데이터셋에 대한 표준편차에 의한 곡선의 아래면적

Table 1. The Area Under Curve(AUC) with Standard Deviation on MNIST dataset.

Normal Class	DCAE	IF	ANOGAN	SVDD	DSVDD	ADGAN (ours)
0	97.6 ± 0.7	98.0 ± 0.3	96.6 ± 1.3	98.6 ± 0.0	98.0 ± 0.7	<b>99.3 ± 0.1</b>
1	98.3 ± 0.6	97.3 ± 0.4	99.2 ± 0.6	99.5 ± 0.0	99.7 ± 0.1	<b>99.8 ± 0.1</b>
2	85.4 ± 2.4	88.6 ± 0.5	85.0 ± 2.9	82.5 ± 0.1	<b>91.7 ± 0.8</b>	90.2 ± 1.2
3	86.7 ± 0.9	89.9 ± 0.4	88.7 ± 2.1	88.1 ± 0.0	91.9 ± 1.5	<b>93.6 ± 0.8</b>
4	86.5 ± 2.0	92.7 ± 0.6	89.4 ± 1.3	94.9 ± 0.0	94.9 ± 0.8	<b>95.6 ± 0.7</b>
5	78.2 ± 2.7	85.5 ± 0.8	88.3 ± 2.9	77.1 ± 0.0	88.5 ± 0.9	<b>91.9 ± 1.9</b>
6	94.6 ± 0.5	95.6 ± 0.3	94.7 ± 2.7	96.5 ± 0.0	<b>98.3 ± 0.5</b>	97.3 ± 0.7
7	92.3 ± 1.0	92.0 ± 0.4	93.5 ± 1.8	93.7 ± 0.0	94.6 ± 0.9	<b>95.4 ± 1.2</b>
8	86.5 ± 1.6	89.9 ± 0.4	84.9 ± 2.1	88.9 ± 0.0	93.9 ± 1.6	<b>95.2 ± 0.4</b>
9	90.4 ± 1.8	93.5 ± 0.3	92.4 ± 1.1	93.1 ± 0.0	<b>96.5 ± 0.3</b>	94.5 ± 1.2

필기체 데이터를 사용한 비정상 검출

필기체 데이터는 비정상 검출 알고리즘의 성능 평가 척도로 사용되는데 현재 최고 성능을 보여주고 있는 방법은 Deep One-Class Classification(L. Ruff et al. 2018)이다. 따라서 본 논문에서는 정확한 성능 검증을 위해 위의 논문에서 제안한 실험 방법을 그대로 차용하여 성능평가를 수행하였다. 필기체 데이터는 0~9까지의 65,000장의 숫자 이미지로 구성되어 있고, 각 클래스 당 6,500장의 이미지가 속해있다. 본 논문은 5,500장의 단일 클래스 이미지(정상 이미지)와 500장의 나머지 클래스 이미지(비정상 이미지)를 ADGAN의 학습을 위해 사용하였고, 성능 검증을 위한 시험 이미지는 각 클래스에서 1,000장씩 추출하여, 총 10,000장의 이미지로 구성하였다. 필기체 데이터가 총 10개의 클래스로 구성되어 있기 때문에, 정상 이미지의 클래스를 바꿔가며 총 10 번의 실험을 수행할 수 있었고, 그 결과는 위의 표와 같다(See Figure 2). 이 때, 각 목적함수를 안정적으로 최적화하기 위해 (6), (7)번 식은 Adam Optimizer<sup>26</sup>, (8) 식은 RMSProp Optimizer<sup>27</sup>를 사용하였다.

검출 및 악화부, 생성부, 검출부의 유용성 검증

본 논문에서 제안하는 비정상 점수의 효율성을 검증하기 위해, 0~4까지의 필기체 숫자 이미지를 정상 이미지, 5~9까지의 필기체 숫자 이미지를 비정상 이미지라고 가정한 후, 다음의 실험을 수행하였다. 20,000장의 정상 이미지를 사용해 제안된 구조를 학습시키고, 1,000장의 정상 이미지와 9,000장의 비정상 이미지를 혼합하여 모델 성능을 평가하였다(See Figure 3).

표 2. 목적함수가 비정상 검출에 적합한지 확인할 수 있는 표  
Table 2. A table to see where objective function is performing well for anomaly detection or not.

$$L = P(X_{ori}, X_{L,D}) + MSE(X_{L,D}, X_{gen})$$

	개선 및 악화부	생성부 추가	생성부, 검출부 추가
목적함수	$L$	$L + L_{cycle}$	$L + L_{cycle}$
AUC	0.880225	0.909272	0.920543

실험 결과, 본 논문에서 제안한  $L_{cycle}$ 와  $L_{Dis}$ 를 적용하면 비정상 검출의 성능을 나타내는 지표인 Area Under Curve(AUC) 값이 각각 0.029047, 0.011271 만큼 증가한다는 것을 수치적으로 확인할 수 있었고, 이는 제안된 목적함수들이 비정상 탐지에 적합하다는 것을 의미한다.

ADGAN을 이용한 용접 결함 검사

공장에서 용접 작업을 할 때, 미용접, 크레이터, 이물질 흡착 등과 같은 용접 결함이 발생할 수 있다. 이러한 결함은 최종 제품에 악영향을 미칠 수 있기 때문에 용접 작업이 끝난 후, 결함 유무를 검사하는 것은 필수적이다.

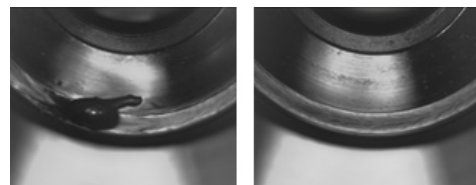


그림 2. 정상 용접 이미지(왼쪽), 비정상(결함이 존재하는) 용접 이미지(오른쪽)

Fig. 2. Images with and without welding defect

용접부 결함 검사는 대표적으로 비전 카메라를 이용한 방법과 초음파를 이용한 방법이 있지만, 본 논문에서는 비전 카메라를 사용하여 촬영한 결함부 사진을 통해 결함의 유무를 판정하고자 한다.

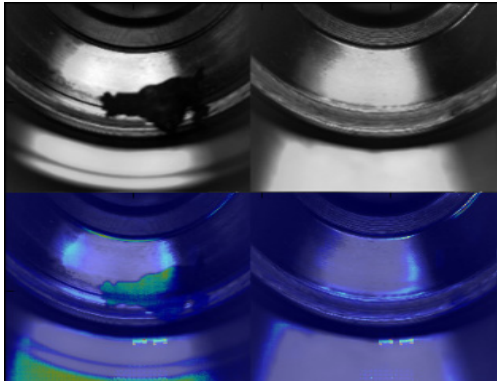


그림 3. 발견된 결함위치의 시각화 왼쪽(결함 존재) 오른쪽(결함 미존재)

Fig. 3. Visualizations of expected defect locations. left column(defect), right column(no defect).

사진으로 구성되어 있는데 제안된 구조의 평가를 위해 정상적인 용접부 사진 3,000장을 사용하여 ADGAN을 학습시켰고, 950장의 정상적인 용접부 사진과 50장의 결함 사진을 혼합하여 성능평가를 진행하였다. 또한, 앞의 이론적 배경에서 설명한 비정상 가시화 방법을 적용해본 결과(See Figure 5) 제안하는 ADGAN이 비지도 학습의 방법론임에도 불구하고 성공적으로 결함의 위치를 파악했음을 시각적으로 확인할 수 있다.

#### IV. 결 론

본 논문은 비지도 학습의 대표 주자인 지각 변분 오토 인코더와 적대적 생성 신경망을 활용하여 노이즈가 심한 입력 이미지에 대해 강건한 비정상 검출을 할 수 있는 ADGAN을 제안하였다. ADGAN은 정상 이미지와 비정상 이미지의 픽셀 단위 거리뿐만 아니라, 이미지의 질적인 면도 같이 고려하기 때문에 기존의 방법론들 보다 더 좋은 검출 성능을 보여주었고, 이를 벤치마크 데이터를 사용하여 실험적으로 증명하였다. 또한, ADGAN을 실제 용접 결함 검사 문제에 적용해보았고, 제안하는 방법이 실제 산업 현장의 결함 검출에도 어느 정도 적용 가능하다는 것을 실험적으로 확인 하였다.

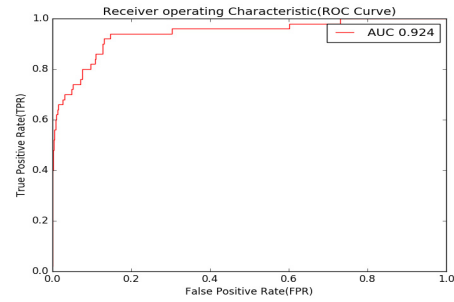


그림 6. 용접 결함 검사에 대한 ROC 곡선과 AUC 값

Fig. 6. ROC curve and AUC value on Welding defect Inspection.

#### Acknowledgement

이 과제는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음

#### References

- [1] C. Callegari, S. Vaton, and M. Pagano. A New Statistical Approach to Network Anomaly Detection. In *Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems(SPECTS)*.2018.
- [2] E. Eskin. Anomaly Detection over Noisy Data using Learned Probability Distribution. In *Proceedings of the International Conference on Machine Learning(ICML)*. pages 255-262, 2000. DOI: <https://doi.org/10.7916/D8C53SKF>
- [3] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger. Densely Connected Convolutional Networks. In *Computer vision and Pattern Recognition(CVPR)*. pages 4700-4708, 2017. arXiv:1608.06993
- [4] S. Zhang, L. Wen, X. Bian, Z. Lei, S. Z. Li. Single-Shot Refinement Neural Network for Object Detection. In *Computer vision and Pattern Recognition(CVPR)*. 2018. DOI: 10.1109/CVPR.2018.00442
- [5] N. Kodali, J. Abernethy, J. Hays, Z. Kira. On Convergence and Stability of GANs. In *arXiv preprint arXiv:1705.07215*. 2017.
- [6] P. Seeböck, S. Waldstein, S. Klimesch, B. S. Gerendas, R. Donner, T. Schlegl, U. Schmidt-Erfurth, G. Langs. Identifying and Categorizing Anomalies in Retinal Imaging Data. In *arXiv preprint arXiv:1612.00686*. 2016.
- [7] T. Schlegl, P. Seeböck, S. M. Waldstein, U.

- Schmidt-Erfurth. Unsupervised Anomaly Detection With Generative Adversarial Networks to Guide Marker Discovery. In *International Conference on Information Processing in Medical Imaging(IPMI)*. pages 146-157, 2017.  
DOI: [https://doi.org/10.1007/978-3-319-59050-9\\_12](https://doi.org/10.1007/978-3-319-59050-9_12)
- [8] M. Sabokrou, M. Khalooei, M. Fathy, E. Adeli. Adversarially Learned One-Class Classifier for Novelty Detection. In *Computer vision and Pattern Recognition(CVPR)*. pages 3379-3388, 2018.  
DOI: 10.1109/CVPR.2018.00356
- [9] L. Ruff, R. A. Vandermeulen, N. Görnitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, M.Kloft. Deep One-Class Classification. In *Proceedings of the International Conference on Machine learning*. pages 4393-4402, 2018.
- [10] A. Vasilev, V. Golkov, M. Meissner, I. Lipp, E. Sgarlata, V. Tomassini, D. K. Jones, D. Cremers. q-Space Novelty Detection with Variational Autoencoders. In *arXiv preprint arXiv:1806.02997*. 2018.
- [10] A. Vasilev, V. Golkov, M. Meissner, I. Lipp, E. Sgarlata, V. Tomassini, D. K. Jones, D. Cremers. q-Space Novelty Detection with Variational Autoencoders. In *arXiv preprint arXiv:1806.02997*.2018.
- [11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B.Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. Generative Adversarial Nets. In *proceedings of Neural Information Processing Systems(NIPS)*. 2014.
- [12] A. Radford, L. Metz, S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In *Proceedings of International Conference on Learning Representations(ICLR)*. 2016.
- [13] X. Hou, L. Shen, K. Sun, G. Qiu. Deep Feature Consistent Variational Autoencoder. In *Winter Conference on Applications of Computer Vision(WACV)*,pages 1133-1141, 2017.
- [14] D. Berthelot, T. Schumm, L. Metz. BEGAN: Boundary Equilibrium Generative Adversarial Networks. In *arXiv preprint arXiv:1703.10717*. 2018.
- [15] D. P. Kingma, M. Welling. Auto-Encoding Variational Bayes. In *Proceedings of the International Conference on Learning Representations(ICLR)*, 2014.
- [16] Y. Blau, T. Michaeli. The Perception-Distortion Tradeoff. In *Conference on Computer Vision and Pattern Recognition(CVPR)*, 2018.
- [17] M. L. Shyu, S. C. Chen, K. Sarinnapakorn, L. Chang. A Novel Anomaly Detection Scheme Based on Principal Component Classifier. In *proceedings of the IEEE Foundations and New Directions of Data Mining Workshop*. pages 172-179, 2003.
- [18] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representation (ICLR)*. 2014.
- [19] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, V. R. Chandrasekhar. Efficient Gan-Based Anomaly Detection. In *International Conference on learning Representations Workshop(ICLR)*. 2018.
- [20] J. Zhao, M. Mathieu, Y. LeCun. Energy-Based Generative Adversarial Networks. In *International Conference on Learning Representations(ICLR)*. 2017.
- [21] M. Arjovsky, S. Chintala, L. Bottou. Wasserstein GAN. In *proceedings of International Conference on Machine Learning(ICML)*. pages 214-223, 2017.
- [22] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, S. P. Smolley. Least Squares Generative Adversarial Networks. In *International Conference on Computer Vision(ICCV)*. pages 2794-2802. 2017.
- [23] J. Y. Zhu, T. Park, P. Isola, A. A. Efros. Unpaired Image-to-Image Translation using Cycle Consistent Adversarial Networks. In *International Conference on Computer Vision(ICCV)*. pages 2223-2232, 2017.
- [24] G. E. Hinton, R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. In *Science* 313, pages 504-507, 2006.
- [25] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin. Improved Training of Wasserstein GANs. In *Advances in Neural Information Processing Systems(NIPS)*. 2017.
- [26] D. P. Kingma, J. L. Ba. ADAM: A method for Stochastic Optimization. In *Proceedings of the International Conference on Learning Representation(ICLR)*. 2015.
- [27] S. Ruder. An Overview of Gradient Descent Optimization Algorithms. In *arXiv preprint arXiv:1609.04747*. 2016.
- [28] S. Ioffe, C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the International Conference on Machine Learning(ICML)*,pages 448-456, 2015.
- [29] In-Kyu Park. A Study on the Prediction of the Nonlinear Chaotic Time Series Using Genetic Algorithm based Fuzzy Neural Network The Journal of The Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 11, No. 4, pp. 91-97, 2011.
- [30] Yun-Kyung Park, Hye-Won Lee, Sang-Moon Kim. Identify the Failure Mode of Weapon System (or equipment) using Machine Learning Journal of the Korea Academia-Industrial cooperation Society(JKAIS), Vol 19, No.8 pp.64-70 2018



## 저 자 소 개

서 태 문(준회원)



- 2019년 : 부산대학교 기계공학과 (학사)
- 2019년~ 현재 : 부산대학교 기계공학과 (석사)

강 민 국(준회원)



- 2019년 : 부산대학교 기계공학과 (학사)
- 2020년~ 현재 : 포항공과대학교 컴퓨터공학과 (석사)

강 동 중(정회원)



- 1988년 : 부산대학교 정밀기계공학 (학사)
- 1990년 : 한국과학기술원 기계공학 (석사)
- 1999년 : 한국과학기술원 자동차 및 설계공학 (박사)
- 2004~2005년 : Cornell University 기계공학과 (박사 후 연수)
- 2006~현재 : 부산대학교 기계공학부 (교수)