RESEARCH

KEPCO Journal on Electric Power and Energy
p-ISSN 2465-8111, e-ISSN 2466-0124
Volume 8, Number 2, December 2022, pp. 137-141
DOI 10.18770/KEPCO.2022.08.02.137

Effective Analsis of GAN based Fake Date for the Deep Learning Model

딥러닝 훈련을 위한 GAN 기반 거짓 영상 분석효과에 대한 연구

Seungmin Jang, Seungwoo Son, Bongsuck Kim
장승민, 손승우, 김봉석

Abstract

To inspect the power facility faults using artificial intelligence, it need that improve the accuracy of the diagnostic model are required. Data augmentation skill using generative adversarial network (GAN) is one of the best ways to improve deep learning performance. GAN model can create realistic-looking fake images using two competitive learning networks such as discriminator and generator. In this study, we intend to verify the effectiveness of virtual data generation technology by including the fake image of power facility generated through GAN in the deep learning training set. The GAN-based fake image was created for damage of LP insulator, and ResNet based normal and defect classification model was developed to verify the effect. Through this, we analyzed the model accuracy according to the ratio of normal and defective training data.

Keywords: Generative Adversarial Network (GAN), Visual intelligence, Fake image, Power facility

I. INTRODUCTION

최근 딥러닝 기반의 이미지 인식 기술과 인지 기술이 크게 발전함에 따라 여러 산업 분야에서 시각 지능을 활용한 결함진단 기술개발이 활발하게 진행 중이다. 전력산업에서도 설비 고장 예방, 자산관리를 위해 인공지능을 활용한 지능형 설비 감시 기술개발을 진행 중이다[1][2].

시각지능을 활용한 설비결함 검출 기술을 개발할 때 가장 어려운 점은 발생 빈도가 적은 결함의 데이터 확보이다. 이러한 문제를 해결하기 위해 현실과 거의 유사한 가상의 데이터를 생성하여 인공지능 학습성능을 높이는 Generative Adversarial Network(GAN) [3] 기반의 전력설비 데이터 증강 기술을 개발하였다.

이에 본 연구에서는 GAN을 통해 생성한 가상의 전력 설비 이미지를 실제 딥러닝 학습에 적용하여,그 효과를 분석하고 가상 데이터 생성 기술의 적용 타당성을 검증하고자 한다. 배전 LP 애자 박리/파손 설비를 대상으로 GAN 기술을 적용하여 가상의 전력 설비 이미지를 생성하였고, 효과 검증을 위해 ResNet 기반의 전력설비 정상 및 결함 분류 모델을 개발하여, 정상과 결함 학습데이터의 비율에 따른 모델 정확도를 분석하였다. 2장에서는 GAN 기반 데이터 증강 모델을 기술하고, 3장에서는 가상 결함 데이터 효과를 검증하기 위해 개발한 정상 및 결함 분류 모델을 소개한다. 4장에서는 실험 방법 및 실험 결과를 정리하였고, 5장에서 결론으로 마무리 한다.

II. GAN 기반 데이터 증강 모델 개발

A. Generative Adversarial Network(GAN) 기술 개요

GAN은 생성 네트워크(Generator)와 판별 네트워크(Discriminator)라고 불리는 두 모델이 동시에 적대적인(adversarial) 과정으로 학습하는 최신 딥러닝 모델이다. 생성자 G는 실제 데이터인 전력설비 데이터의 분포를 학습하고, 판별자는 원래의 전력설비 데이터인지, 아니면 인공지능 모델을 통해서 생성된 데이터인지 구분한다. 생성자의 학습 과정은 가상의 전력설비 이미지를 잘 생성해서 판별자를 속일 확률을 높이고, 반대로 판별자 는 제대로 구별해내는 확률을

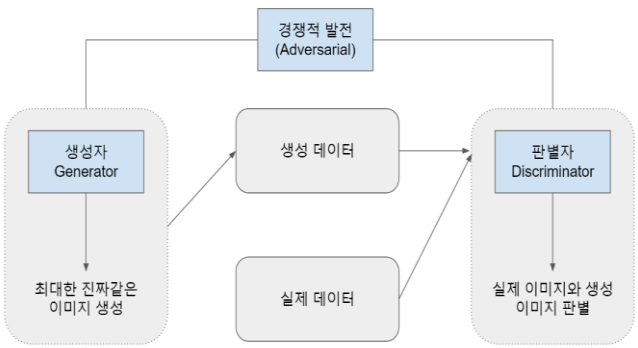


Fig. 1. Generative Adversarial Network 개념

Article Information

Manuscript Received August 17, 2022, Accepted September 16, 2022, Published online December 30, 2022

The authors are with KEPCO Research Institute, Korea Electric Power Corporation, 105 Munji-ro Yuseong-gu, Daejeon 34056, Republic of Korea.

Correspondence Author: *Jintae Cho (jintae.cho@kepco.co.kr)*

높이는 학습 과정이라고 볼 수 있다.

생성자의 인풋은 랜덤 벡터(노이즈)로, 가장 초기의 아웃풋도 노이즈 형태이다. 이는 시간을 거듭하면서 판별자의 피드백을 받으며 생성자는 좀 더 사실적인 이미지를 생성하는 법을 학습하게 된다. 판별자 역시 생성된 이미지와 실제 이미지를 비교하며 고도화 작업을 거쳐 생성자로 하여금 판별자를 속이기 어렵게 만든다.

B. GAN 기반 전력설비 가상 결함 이미지 생성 모델 소개

GAN 모델은 생성자와 판별자의 구성과 사용 목적, 모델링 형태에 따라 다양하다. 본 연구에서는 데이터 셋이 소량일 때 성능을 낼 수 있는 유용한 모델인 StyleGan2-Ada[4]를 이용하였다. StyleGAN2-ADA는 Data Augmentation 기능을 추가하여, 판별 네트워크의 입력 이미지를 증강시켜 학습 데이터가 부족한 문제를 어느 정도 해결하게 해준다. 이러한 원리로 적은 수의 전력 설비 학습 데이터라고 해도 학습이 가능하다.

C. 학습 데이터 구성 및 모델 개발 환경

1) 학습용 데이터 분석 및 재구성

수집한 원본 데이터에 대하여 불필요하거나 의심스러운 데이터들은 이미지 수정 도구로 결함 부분을 지우거나 학습 데이터에서 제외하였다. 또한, LP애자 주변부에 심한 방해물들이 있는 이미지들 또한 학습에서 제외했다. 기본적으로 원본 데이터셋 수가 아주 적은 만큼 최대한 좋은 품질의 데이터만 학습에 포함했다

2) 데이터 학습 환경

StyleGAN2-ADA 코드 구현은 리눅스 우분투(20.04) 환경에서 진행하였고 원활한 실험을 위해 GPU는 48GB 메모리의 NVIDIA QUADRO RTX 8000을 4개 사용하였다. 딥러닝 프레임워크는Python: PyTorch: 1.7 및 CUDA toolkit: 11.0을 사용했으며, Python 3.7 언어 기반으로 구현하였다.

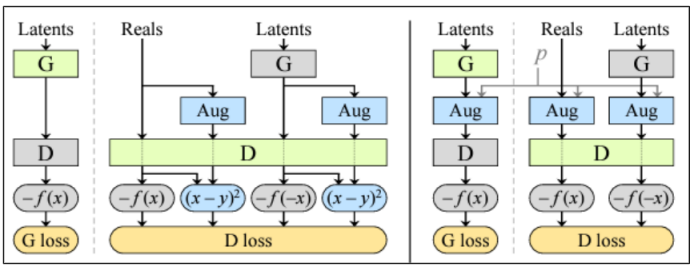
3) 생성모델 학습 – 학습 하이퍼 파라미터(hyper parameters)

모든 학습은 ‘seed=0’으로 고정해서 진행하였다. Seed 값을 고정은 인풋으로 들어오는 난수를 예측 가능하게 한다는 것을 의미한다. 실험 할 때에는 알고리즘 성능에 따라 지표가 개선이 된 건지, 아니면 random 성이 뛰어나서 지표가 개선된 것인지 확실히 하는 것이 필요한데, seed값을 고정함으로써 특정 변수에 의한 변환된 결과 분석이 가능하다.

초기 학습 시에는 원본 데이터의 수량이 적어 속도는 느리지만 정밀한 학습을 위해 batch size를 ‘4’로 세팅하였고, 어느 정도 학습이 이루어진 이후로는 batch size를 16으로 조정하여 학습을 수행하였다.

4) LP 애자 이미지 생성

가상의 이미지 생성에 있어서 어려운 점 중에 하나는 학습데이터에 없었던 영역을 생성해야 하는 것이다. 하지만 생성자는 빈도가 적은 특정들은 학습이 어렵고 이미지도 생성하기 어렵다. 그 대신 나쁜 품질의 이미지를 생성하게 된다. 이와 같은 문제를 피하고자, StyleGAN은 중간벡터(w)를 잘라내어(truncation) ‘평균적인’ 중간 벡터에 가까이 가도록 강제했다. 즉, 모델을 학습한 후, 평균적인 w



(a) previous work (b) StyleGAN2-ADA

Fig.2. (a) 기존 bCR(balanced consistency regularization) 기법.
(b) StyleGAN2-ADA 기법

TABLE 1
학습용 데이터 구성 현황

구 분	LP애자_정상	LP애자_박리파손
수 량	1,739장	576장

Fig.3. 생성 이미지 결과 샘플 (LP 애자 박리파손)

는 많은 인풋 벡터를 선택해 생성되는데, 매핑 네트워크를 통하여 w를 생성하고 이들의 평균을 구한다. 새 이미지를 생성할 때는 매핑 네트워크를 활용하지 않고, w의 평균값으로 변환해 사용한다. 0과 1사이의 값을 지니는 truncation을 낮게 조정할수록 생성 이미지의 다양성을 보장하고, 높게 조정할수록 학습 데이터의 분포를 생성하려는 경향이 있어서, 적절한 값을 선택하여 생성하는 것이 중요하다.

III. 애자 정상 및 결함 분류모델 개발

A. 분류모델 선정

GAN으로 결함 데이터를 생성하는 것이 분류 모델의 성능을 높이는 데 효과적인지 파악하기 위해서 현재 확보한 데이터만을 사용해서 기준이 되는(이하, 베이스라인) 모델을 만들어야 할 필요가 있다. 결함 데이터의 분류는 자연 이미지 분류에 관한 과제에서 뛰어난 성능을 내는 것으로 알려진 ResNet(레스넷) 계열의 딥러닝 모델을 사용하였다.

일반적으로 딥러닝 모델은 레이어의 깊이가 성능에 큰 영향을 끼친다. 그러나, 레이어가 너무 깊어지면 그라디언트가 소실되는 문제가 발생하게 되고, 역전파가 되어도 앞부분의 레이어 일수록 가중치에 끼치는 영향이 0으로 수렴해, 더 이상 학습이 진행되기 어려워 모델의 성능이 향상되지 않는 문제가 있다. ResNet은 이 문제

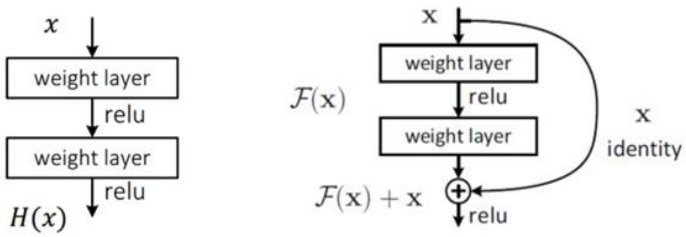


Fig.4. 일반적인 딥러닝 아키텍쳐의 레이어(좌)와 ResNet의 잔차 학습 방식(우)

TABLE 2
ResNet-18 분류 모델에 사용한 설정

설정 파라미터	내용
epochs	학습 루프의 반복 횟수
learning_rate	학습률
batch_size	(train 또는 test의) 배치 사이즈
valid_size	학습 데이터 중 평가 데이터로 사용할 데이터의 비율
train_path, test_path	학습 또는 테스트 데이터의 경로
base_model	ResNet 계열의 모델명

를 skip-connection을 활용한 잔차 학습(residual learning)을 통해 아주 깊은 모델에서도 그라디언트가 소실되지 않도록 문제를 해결한 아키텍처이다.

ResNet은 레이어의 깊이에 따라 18, 34, 50, 101 그리고 152까지 다양하게 사용되고 있으나, 1,000개의 사물을 분류하는 이미지넷-1k 분류 챌린지에서 ResNet-18을 특징추출기로 활용한 최근 사례들의 top-1 accuracy가 대부분 70~72% 정도로 나쁘지 않은 성능을 보인 점과 파라미터 수도 11M으로 다른 모델에 비해 매우 적은 편임을 고려할 때, 정상과 결함만을 분류하는 본 연구에서는 학습 속도나 성능 면에서 가장 적절하다고 판단이 드는 ResNet-18을 적용하였다.

학습으로 생성되는 모델의 가중치 파일은 valid_size에 따라 가변적이긴 하나, 일반적인 기계학습 과제에서 이뤄지는 데이터 분할 비율에 따라 30%(0.3)로 고정하고, 이렇게 분할된 데이터로 매 epoch마다 평가 정확도를 측정해 이전 epoch보다 향상되었을 시에만 저장하도록 구현했다. 학습은 총 6 epoch 만을 실행했고, 그 이상 학습을 진행하면 손실 값이나 정확도 등으로 미루어 볼 때 과적합(overfitting)이 일어나기 쉽다는 사실이 실험적으로 확인되었다. 본 연구의 경우 epoch 수를 6으로 고정한 모델을 베이스라인으로 설정하고, 추후 GAN으로 생성된 이미지가 추가되었을 때에도 정확한 모델 비교를 위해 학습은 6 epoch까지로 고정하였다. 또한, GAN으로 생성된 데이터를 추가했을 때 모델의 성능이 향상되는지 확인하는 것이 본 과제의 주된 목적이므로, 베이스라인 모델을 학습할 때는 여타 데이터 증강 기술은 사용하지 않았다.

IV. 실험 결과

A. 실험방법 및 절차

실제 현장에서 취득한 LP 애자 박리/파손 이미지 576장에 추가로 가상의 박리파손 이미지 1,163장을 GAN 생성기로부터 추출하면, 총 1,739장의 이미지가 확보되어 정상 애자 이미지 수량과 정확히 1:1의 비율로 데이터셋을 구성할 수 있다. 그러나 분류 성능 변화의 정확한 추이를 살펴보기 위해 정상-결함 이미지 구성 비율별로 아래 TABLE 3과 같이 데이터셋을 상세하게 구분한 뒤 모델을 학습하였다.

TABLE 3
GAN으로 생성한 LP애자 결함(박리파손) 이미지 수량에 따른 데이터셋 구성

구 분	정상 이미지	결함 이미지 (실제결함+가상결함)	비율	비 고
데이터셋 1	1,739장	576장 (576장+0장)	1:0.3	
데이터셋 2	1,739장	960장 (576장+384장)	1:0.5	
데이터셋 3	1,739장	1,344장 (576장+768장)	1:0.8	이전 384장을 모두 포함
데이터셋 4	1,739장	1,739장 (576+1,163)	1:1	이전 768장을 모두 포함

TABLE 4
학습에 사용한 하이퍼 파라미터(모델 1~4 공통 적용)

설정 파라미터		내 용
epochs	6	학습 루프의 반복 횟수
learning_rate	0.001	SGD 최적화 함수의 학습률
batch_size	4	(train 또는 test의) 배치 사이즈
valid_size	30%	학습 데이터 중 평가 데이터로 사용할 데이터의 비율
base_model	ResNet-18	ResNet 계열의 모델명

TABLE 5
혼동행렬(Confusion matrix)의 정의

구 분	정상으로 예측	비정상으로 예측
실제로 정상	True Positive (TP) =정상을 정상으로 예측	False Negative (FN) =정상을 비정상으로 잘못 예측
실제로 비정상	False Positive (FP) =비정상을 정상으로 잘못 예측	True Negative (TN) =비정상을 비정상으로 예측

데이터셋을 제외한 모델의 학습 환경, 모델 종류 및 하이퍼 파라미터는 베이스라인 모델을 학습한 조건과 모두 같게 설정했다 (TABLE 4). 이하, 편의를 위해 각 데이터셋의 번호에 맞춰 학습한 모델을 모델 1~4'로 명명하였다.

'모델 2~4'는 동일한 시드값을 설정해서 미니 배치를 일관성 있게 구성하도록 학습했고, 총 2세트의 실험을 거쳤다. 각 세트는 모든 변수가 같으나 단지 초기 시드값만 다르게 학습하였다. 그리고, 모든 모델은 베이스라인 모델에서 사용한 테스트셋(정상, 박리/파손 데이터 각100장)으로 공평하게 성능을 테스트했다. 모델 1(베이스라인 모델)에서 사용했던 것과 동일한 테스트셋을 사용해 '모델 2~4'의 성능을 측정했다.

B. 성능지표 정의

분류 모델의 성능은 항상 혼동행렬(Confusion matrix)과 ROC

(Reciever Operating Characteristics) 곡선을 이용해 평가한다. ROC 곡선이란, 분류 상황에서 클래스 결정 기준에 대한 진단 능력을 시각화한 것이라 생각할 수 있다.

딥러닝 모델은 혼동행렬을 기준으로, 다양한 관점에서 평가할 수 있다. 먼저 정확도(또는 정답률, accuracy)는 전체 테스트 샘플 중에서 정확하게 예측한 샘플 수의 비율을 의미하며, 식 1과 같다.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

다음으로 정밀도(precision)는 정상(Positive) 클래스에 속한다고 예측한 샘플 중 실제로 정상(Positive) 클래스에 속하는 샘플 수의 비율을 말하며 일반적으로 높을수록 좋은 모형이라 할 수 있다.

$$precision = \frac{TP}{TP+FP} \quad (2)$$

재현율(recall)은 실제 정상(Positive) 클래스에 속한 샘플 중에서 정상(Positive) 클래스에 속한다고 예측한 샘플 수의 비율을 뜻하고, 이 역시 일반적으로 높을수록 좋은 모형이라 평가한다. 재현율은 민감도(Sensitivity) 또는 TPR(True Positive Rate) 로도 불린다.

$$recall = \frac{TP}{TP+FN} \quad (3)$$

이에 반해 위양성률(fall-out)은 낮을수록 좋은 모형이라 평가하는데, 실제 비정상 클래스에 속하지 않는 샘플 중에 비정상 클래스에 속한다고 예측한 샘플의 비율을 일컫는다. 위양성률 역시 FPR (False Positive Rate)로 부르는 경우가 있으며, 경우에 따라서는 1에서 FPR을 뺀 값인 특이도(Specificity)로 모델을 평가하기도 한다.

$$fall-out = \frac{FP}{FP+TN} \quad (4)$$

위에서 소개한 지표 중 재현율과 위양성률은 대개 양의 상관성을 띈다. 재현율을 높이려면 비정상을 판단하는 기준, 즉 역치(threshold)를 낮춰 샘플을 비정상으로 판단하기 쉽게 하면 된다. 그러나, 이로 인해 정상임에도 불구하고 비정상으로 판단되는 샘플 데이터가 증가하기 때문에 위양성률역시 동시에 증가한다. 반대로, 역치를 높여 위양성률을 낮추게 되면 비정상임에도 정상으로 판별되는 샘플이 증가하기 때문에 재현율 역시 떨어진다. 이와 같이 클래스 판별의 기준이 되는 역치의 변화에 따른 위양성률과 재현율의 변화를 시각화한 것이 ROC 곡선이다. Fig.5는 ROC 곡선이 그려지는 방식을 시각화한 것으로, 역치가 -∞에서 +∞까지, 이동하면서 계산된 TPR을 y축에, FPR을 x축에 그린 것이며, 빨간색의 분포는 정상인 데이터들의 분포, 초록색은 비정상 데이터들의 분포를 나타낸다.

ROC 곡선의 밑면적을 의미하는 AUC(Area Under Curve) 값은 여러 모델의 ROC 곡선이 있을 때, 한눈에 모델의 성능을 비교하기에 적절한 지표다. Fig. 5의 오른쪽의 TPR-FPR 그래프는 전체 면적은 1이고, AUC는 1보다는 작은 값을 가진다. 예를 들어, AUC가 1이라면 ROC 커브는 정확히 사각형 모양을 띠게 되고, 아주 이상적인 분류 모델이 완성되었음을 의미한다.

ROC 곡선의 밑면적인 AUC(Area Under Curve) 값은 여러 모델의 ROC 곡선이 있을 때, 한눈에 모델의 성능을 비교하기에 적절한 지표다. Fig.5의 오른쪽 TPR-FPR 그래프는 전체 면적은 1이고, AUC는 1보다는 작은 값을 가진다. 예를 들어, AUC가 1이라면 ROC 커브는 정확히 사각형 모양을 띠게 되고, 아주 이상적인 분류 모델이 완성되었음을 의미한다.

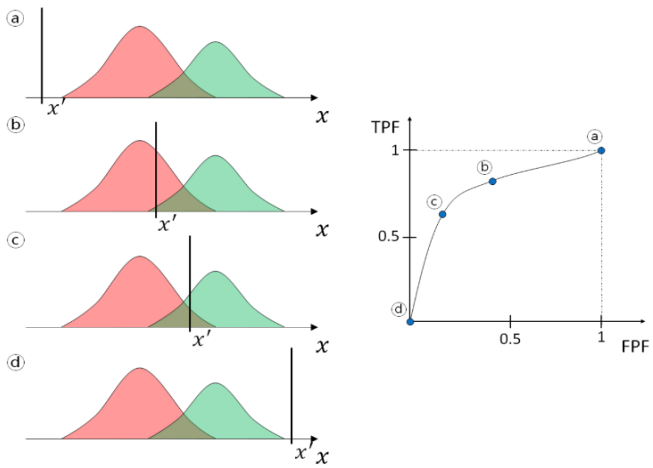


Fig.5. ROC 곡선을 그리는 과정

C. 실험결과

Fig. 6의 혼동행렬과 ROC 곡선은 위에서부터 차례로 '모델1~4'에 대한 테스트 결과다. 각 모델의 정확도와 AUC(ROC 곡선의 밑면적)값을 정리하여 TABLE 6에 정리하였다.

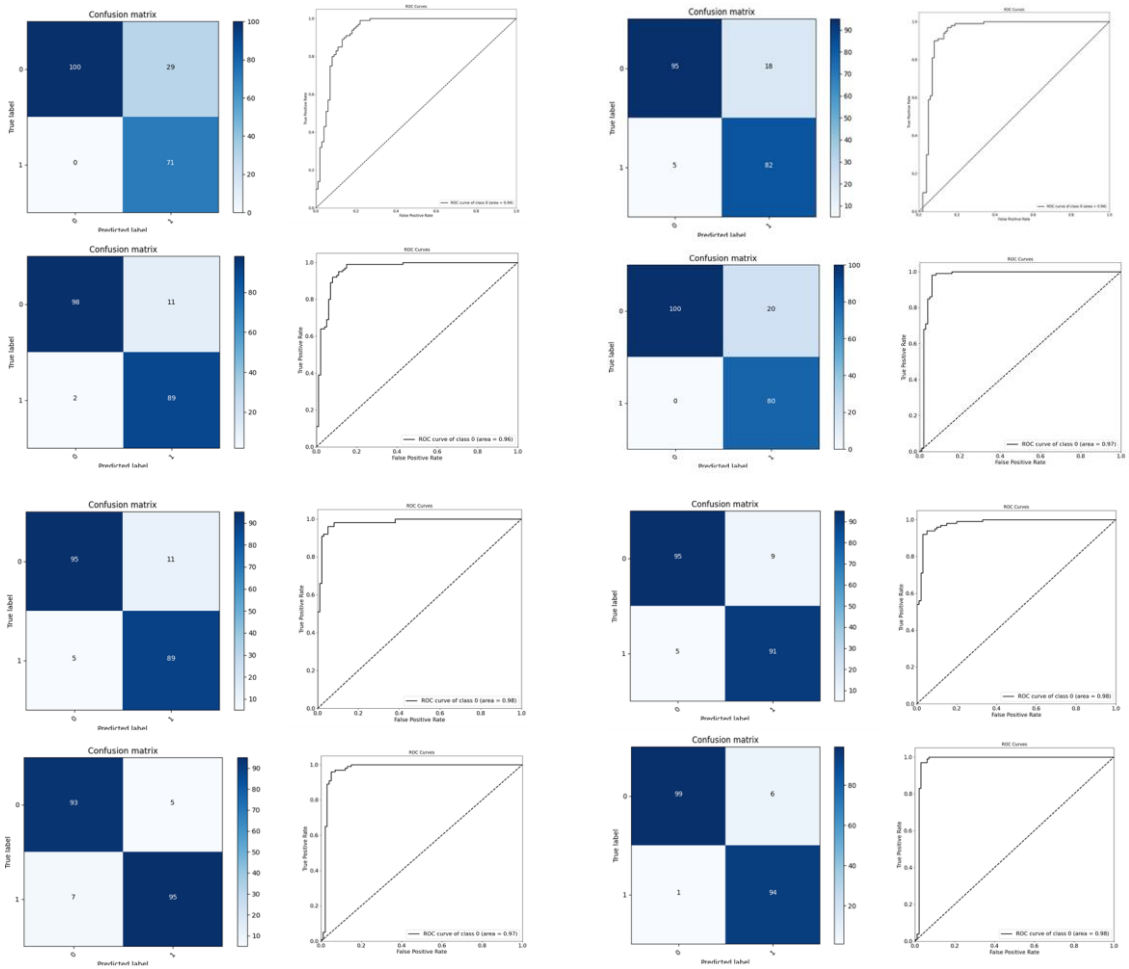
실험결과에서 알 수 있듯이 GAN으로 가상의 박리파손 이미지를 생성 후 데이터셋에 포함시키면, 베이스라인 모델에 비해 정확도와 AUC 값이 크게 향상됨을 확인할 수 있다. 정확도는 두 세트의 실험에서 모두 데이터셋 4를 활용한 학습, 즉, 박리/파손 이미지를 정상 이미지 개수만큼 늘려서(즉, 1:1의 비율) 학습한 경우가 가장 뛰어났다.

Fig. 6-(a) '모델 1'과 '모델 4'의 혼동행렬을 구체적으로 비교해 보면, 정상을 정상으로 예측한 정답수는 변함이 없었으나, 박리/파손을 박리/파손이라고 올바르게 예측한 정답수가 100장 중 71장에서 95장으로 매우 큰 폭으로 증가해, 정확도 향상에 크게 기여 했다. 마찬가지로, Fig. 6-(b)의 '모델 1'과 '모델 4'의 혼동행렬과도 비교해 보면, 이번에는 정상에 관해서는 95장에서 99장으로, 박리파손에 관해서는 82장에서 94장으로 가장 좋은 정확도를 보였다.

AUC 값의 관점에서는, 학습 데이터셋의 정상과 박리/파손의 비율이 1:0.8 정도였던 '모델 3' 역시 '모델 4'만큼이나 뛰어난 성능을 보였다. 이 부분을 현 시점에서 유의미하게 해석하기는 어려우나, 추후에 정상과 박리파손 이미지가 현재보다 다양한 분포를 가진다면, 다양한 실험을 통해 세밀한 추이 분석이 가능할 것으로 보인다.

V. CONCLUSION

본 논문에서는 GAN으로 생성한 가상의 결함 데이터가 실제 모델 정확도 개선 효과가 있는지를 확인하기 위해 ResNet-18 기반의 전력 설비 정상 및 결함 분류모델을 개발하고 학습 데이터셋에 포함된 가상 데이터 비율에 따른 분류모델 정확도 비교하였다. 실험결과 결함 이미지가 정상 이미지에 비해 수량이 적은 조건에서 모



(a) 1세트 (b) 2세트

Fig.6. 모델1~4 (위에서부터 1,2,3,4)의 혼동행렬(좌)와 ROC 곡선(우).

TABLE 6
GAN으로 생성한 박리파손 이미지를 추가한
데이터셋에 따른 모델의 성능 변화

구 분	모델 종류	학습 데이터셋	정확도(%)	AUC	비고
1세트	모델 1	데이터셋 1	85.5	0.94	베이스라인 모델
	모델 2	데이터셋 2	92.0	0.96	
	모델 3	데이터셋 3	93.5	0.98	
	모델 4	데이터셋 4	94.0	0.97	
2세트	모델 1	데이터셋 1	88.5	0.94	베이스라인 모델
	모델 2	데이터셋 2	90.0	0.97	
	모델 3	데이터셋 3	93.0	0.98	
	모델 4	데이터셋 4	96.5	0.98	

델개발을 하는 것보다 가상의 결함 데이터를 증강하여 학습에 포함하는 게 성능이나 정확도가 향상됨을 확인하였다. 본연구결과로 미루어볼 때 결함 이미지 외에도 정상 이미지도 같이 증강하여 정상-결함 이미지 비율을 1:1로 유지하며 10k 이상의 데이터셋을 확보한다면 더욱 완성도 높은 딥러닝 분류모델 개발이 가능할 것으로 생각된다. 향후 정상, 박리/파손 분류뿐 아니라 균열, 아크등 다른 유형의 결함을 동시에 분류하는 다중(Multi-Class) 분류모델에서도 가상 데이터의 학습데이터 활용이 유의미한 기술인지 검증할 계획이다.

ACKNOWLEDGEMENT

This research was supported by Korea Electric Power Corporation under Grant R21IA02.

References

[1] 이동엽, 김준오, 윤정용, 신동열, 최민희(2017). 인공지능 기반 배전 폴리머 현수애자 진단 장비 개발. 대한전기학회 학술대회 논문집, 1547-1548.

[2] 이동엽, 김준오, 윤정용, 최민희(2017). Deep Learning을 활용한 배전 기자재 영상 진단 알고리즘 개발. 대한전기학회 학술대회 논문집, 1504-1505.

[3] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Adv Neural Inf Process Syst 2014:2672-2680.

[4] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training Generative Adversarial Networks with Limited Data. arXiv:2006.06676v2 [cs.CV] 7 Oct 2020.