논문 2022-2-25 http://dx.doi.org/10.29056/jsav.2022.12.25

산업 환경의 객체 검출 성능 개선을 위한 GAN 기반 이미지 데이터 증강

백문기*, 김계경**†

GAN-based Image Data Augmentation for Improving Object Detection Performance in Industrial Environments

Moon-Ki Back*, Kye-Kyung Kim**†

요 약

객체 검출은 예상치 못한 위험 상황을 회피할 수 있도록 작업자에게 자동으로 경보를 제공할 수 있는 중요한 산업 안전 기술 중 하나이다. 그러나 딥 러닝 기반 객체 감지 모델은 더 높은 성능을 달성하기 위해 대규모의 학습 데이터가 요구되며, 데이터 수집 및 레이블링 작업은 인력이 필요한 힘든 작업이다. 이러한 한계를 해결하기 위해, 본 논문에서는 본래의 데이터세트를 더욱 다양한 데이터로 보충할 수 있는 GAN 기반 데이터 증강을 제안한다. 또한, 생성된 데이터의 충실도를 향상시키기 위한 트랜스포머 기반 생성자 네트워크를 제시하며, 비교 연구를 위해 다양한 증강 설정에서 훈련된 기존 객체 감지 모델(YOLOv5)을 평가한다. 평가 결과는 20% 증강된 데이터로 학습한 모델이 객체 특정 성능의 손실 없이 0.9%의 분류 능력이 향상되었음을 보여준다.

Abstract

Object detection is one of the important industrial safety technologies that can automatically provide a worker with alerts to avoid unexpected near misses. However, deep learning-based object detection models require large amounts of training data to achieve higher performance, and data collection and labeling work is laborious and requires human resources. To address these limitations, we propose a GAN-based data augmentation that can supplement the original dataset with more diverse examples. In addition, we present a transformer-based generator network to improve the fidelity of generated data and evaluate the existing object detection model(YOLOv5) trained under different augmentation settings for a comparison study. The evaluation results show that the classification ability of the model trained with 20% augmented data has improved by 0.9% without localization performance losses.

한글키워드 : 생성적 적대 신경망, 데이터 증강, 객체 검출, 딥러닝, 산업 환경

keywords : Generative Adversarial Networks, Data Augmentation, Object Detection, Deep Learning, Industrial Environment

* 한국전자통신연구원 박사후연구원
** 한국전자통신연구원 책임연구원
† 교신저자: 김계경(email: kyekyung@etri.re.kr)
접수일자: 2022.11.30. 심사완료: 2022.12.06.
게재확정: 2022.12.20.

1. 서 론

딥러닝(Deep Learning)을 활용한 객체 검출

(Object Detection) 기술이 점점 고도화되면서 생산된 제품의 불량을 찾아내는 작업, 대규모 물류를 처리하는 로봇, 복잡한 환경을 자율주행하는 지게차에 이르기까지 산업 분야 곳곳에 핵심기술로 자리 잡고 있다. 게다가 산업현장에 만연한 산업 재해를 저감시키기 위해 각종 위험 객체나 작업 행동을 자동으로 검출하는 연구개발도 활성화되고 있다. 건설 현장의 경우 근로자의 부상 위험을 줄이기 위하여 안전모, 안전조끼 착용 여부를 검출하는 딥러닝 모델[1, 2]을 활용할 수 있으며, 제조 및 물류 현장의 경우 중장비와 근로자의 충돌 사고를 예방하기 위해 실시간으로 보행자를 검출하는 딥러닝 모델[3, 4]을 고려해 볼 수 있다.

일반적으로 딥러닝 기반의 모델은 주어진 학습 데이터의 양에 비례하여 성능이 향상된다고 알려져 있다[5, 6]. 딥러닝 기반 객체 검출 분야 역시 충분한 양의 학습 데이터가 필요하며[7], 학습 데이터의 양이 증가할수록 위치 식별(localiza-tion) 및 분류(classification) 성능이 향상된다. 반면, 학습 데이터의 양이 적다는 것은 포함된 객체의 다양성(variety)이 낮다고 해석될 수 있으며, 이를 학습한 딥러닝 기반 객체 검출 모델은 낮은 다양성에 과적합(overfitting)되어 실제 환경(real-world)에서 올바르게 동작하지 않을 가능성이 높다. 데이터 증강(augmentation)은 과적합 문제를 개선하기 위해 학습 데이터를 사용해 학습 데이터의 규모와 다양성을 높이는 기법으로 딥러닝 여러 분야에서 흔하게 사용한다. 예를 들면, 이미지의 일부를 잘라내거나(cropping), 회전(rotation), 이동(translation)하는 등의 데이터 조작(manipulation)[7]은 여러 딥러닝 연구와 챌린지[8]를 통해 과적합 문제를 개선하는데 효과가 있음이 검증되었다.

데이터 조작과 다르게 딥러닝 모델을 사용하여 학습 데이터를 증강하는 접근도 있다. 대표적으로 생성적 적대 신경망(Generative Adversarial Network, GAN)[9]은 학습 데이터의 확률 분포를 학습하고 학습된 분포로부터 새로운 데이터를 생성한다. 이러한 접근은 학습 데이터에 누락된 모집단의 일부를 GAN을 통해 근사함으로써 학습 데이터와 유사한 특징을 가지면서도 전통적인 데이터 조작으로는 생성할 수 없는 새로운 데이터를 생성할 수 있는 장점이 있다. [10, 11]은 GAN을 통해 가상의 이미지를 생성한 연구로, 실제와 가상의 이미지를 분별하기 어려울 정도로 자연스러운(realistic) 이미지 생성이 가능함을 보였다. 게다가 GAN은 의료 분야와 같은 데이터 수집과 활용에 제약이 많은 분야에서 활용 가치가 높다. [12]에서는 CT(Computed Tomography) 이미지 상에 나타난 간 병변(liver lesion) 여부를 분류하는 딥러닝 모델의 성능을 높이기 위하여 GAN 기반으로 CT 이미지를 증강하였다.

본 논문은 객체의 위치 식별 및 분류를 수행하는 딥러닝 객체 검출 모델의 성능을 개선하기 위한 GAN 기반의 이미지 데이터 증강에 대하여 다룬다. 기존의 GAN 기반 이미지 데이터 생성은 이미지 전체(scene)를 생성하기에 이미지 분류에는 직접 사용할 수 있지만, 이미지 내에 존재하는 객체들의 바운딩 박스(bounding box) 정보까지는 생성하지 못하므로 생성된 이미지를 객체 검출에 활용하긴 어렵다. 본 논문에서는 산업현장에서 매우 중요한 객체인 '사람'에 초점을 맞추어, 주어진 학습 데이터로부터 가상의 사람 인스턴스와 바운딩 박스를 생성하는 GAN 기반 데이터 증강을 제안한다. 그리고 제안하는 데이터 증강이 기존 딥러닝 객체 검출 모델에 미치는 영향력을 평가한다.

본 논문은 다음과 같이 구성된다. 2장에서는 기존 딥러닝 모델의 성능을 개선하기 위해 GAN을 사용해 데이터를 증강하는 연구를 살펴보고, 3장에서는 딥러닝 객체 검출의 성능을 향상시키

기 위한 GAN 기반의 이미지 데이터 증강 방법을 제안한다. 4장에서는 본 논문에서 제안한 데이터 증강이 딥러닝 기반 객체 검출 모델에 미치는 영향력을 평가하며, 5장에서는 결론 및 향후 연구에 대하여 논한다.

2. 관련 연구

GAN은 여러 생성 모델(generative model) 중 하나로서 생성자(generator) 네트워크와 판별자(discriminator) 네트워크를 적대적(adversarial)으로 배치하여 학습시키는 독특한 구조를 가진다. 생성자 네트워크는 잠재 벡터(latent vector)로부터 가상의 데이터를 생성하고, 판별자 네트워크는 가상의 데이터와 학습 데이터(실제 데이터)를 구분함으로써 생성자 네트워크에게 피드백을 전달한다. 그리고 이러한 학습 과정을 통해 생성된 데이터가 학습 데이터와 유사한 확률 분포로 근사되는 것을 목표로 한다. 그래서 근사된 확률 분포로부터 생성된 데이터는 학습 데이터와 매우 유사한 특징을 가지면서도 전통적인 데이터 조작으로는 만들어내기 어려운 변이(variation)를 보이는 특성이 있다. 이러한 GAN의 특성은 학습 데이터의 클래스 불균형(imbalance)과 같은 데이터 부족 문제를 해소함으로써 판별 모델(discriminative model)의 과적합 문제를 개선할 수 있다. 게다가 학습이 완료된 GAN은 생성자 네트워크의 가중치(weight)만 사용하여 가상의 데이터를 생성하기 때문에 본래 학습 데이터를 익명화시키는 부가 효과도 얻을 수 있다. 따라서 의료 분야의 개인정보뿐만 아니라 기업 비밀이 보장되어야 하는 산업 분야까지 다양한 목적으로 활용할 수 있다.

[13]은 합성곱 신경망(Convolutional Neural Network, CNN) 구조의 GAN을 사용하여 가상의 흉부 X-ray 이미지를 생성함으로써 학습 데이터의 클래스 불균형 문제를 개선하였다. 그 결과 증강된 X-ray 이미지를 사용하여 학습시킨 판별 모델의 질병 분류 성능이 기존 70%에서 92%로 크게 향상되었다. 이러한 결과는 증강된 데이터를 사용해 판별 모델의 결정 경계(decision boundary)가 특정 클래스에 과적합되는 것을 막았다고 해석할 수 있다. 다만, 전통적인 이미지 조작과의 성능 비교가 제시되지 못했으며, GAN을 통해 생성된 X-ray 이미지의 충실도(fidelity)가 정량적으로 제시되어 있지 않기 때문에 분류 성능이 향상된 이유가 GAN 기반의 데이터 증강이라고 그대로 해석하기엔 무리가 있다.

[14]는 리치(lychee)의 결함(defect) 여부를 검출하는 딥러닝 기반 객체 검출 모델 및 검출 성능 개선을 위한 GAN 기반 데이터 증강을 다루었다. 이 연구에서는 TransGAN[15]을 사용하여 부족한 학습 데이터를 보완하였다. TransGAN은 비교적 최근 등장한 GAN 아키텍쳐로, CNN이 주를 이루는 생성자 네트워크의 구조를 트랜스포머(Transformer)로 변경하여 높은 충실도의 이미지 데이터를 생성했다는 것이 주요 기여이다. 다만, 트랜스포머는 CNN보다 더 많은 학습 데이터를 요구하는 경향이 있는데, TransGAN의 저자들은 DiffAugment[16]를 차용하여 학습 과정에 판별자 네트워크가 과적합되지 않으면서 동시에 생성자 네트워크가 다양성 높은 데이터를 생성하도록 하였다. 따라서 [14]의 저자들은 매우 적은 양의 리치 이미지 데이터를 가지고도 GAN 기반의 데이터 증강이 가능했던 것으로 보이며, 증강된 데이터를 사용해 기존 딥러닝 검출 모델의 성능을 약 1% 향상시켰다.

[17]에서는 기존 보행자 검출 모델의 야간 검출 성능을 개선하기 위해 SamplePairing[18] 기반의 SeAFusion 알고리즘을 제안하여 학습 데이터(이미지)의 조도를 높였으며, ResNet[19] 아키

텍처의 GAN을 사용하여 객체(보행자) 인스턴스의 다양성을 높였다. 그리고 제안한 SeAFusion과 GAN 기반 데이터 증강이 객체 검출 성능에 어떠한 영향력을 미치는지 YOLOv5[20]를 사용하여 검출 성능을 평가하였다. SeAFusion을 적용하였을 때 mAP(mean Average Precision)는 0.209에서 0.795로 크게 향상되었으며, GAN 기반 데이터 증강을 추가적으로 적용하여 0.854까지 끌어올렸다. 이처럼 보행자 검출 성능이 크게 향상된 이유는 FN(False Negative, 정답은 보행자이나 검출기는 보행자로 분류하지 못함)의 비율이 크게 감소하였기 때문으로 분석되었다.

3. GAN 기반 이미지 데이터 증강

3.1 데이터세트

표 1. 산업 환경의 객체 검출을 위한 데이터세트
Table 1. The dataset for detecting objects in industrial environment

Class ID: name	Number of instances		
	Train	Validation	Test
0: person	78,505	10,034	10,115
1: forklift	53,099	6,844	7,000
2: face	69,920	9,108	9,291
3: upper body	91,139	11,685	11,900
4: lower body	50,184	6,425	6,496
5: battery car	3,453	1,083	1,094

본 논문에서는 다양한 객체들이 존재하는 실제 산업현장을 주행하는 지게차로부터 약 15만장의 이미지 데이터를 입수하였다. 그리고 해당 산업현장의 안전관리자 의견을 반영하여 6가지의 주요 객체(사람, 지게차, 얼굴, 상체, 하체, 배터리카)를 정의하였다. 사람의 경우 산업 환경의 각종 구조물들로 인해 전신의 일부가 가려지는 상황이 많기 때문에 전신 뿐만 아니라 얼굴, 상체, 하체로도 구분하여 레이블링(labeling)을 수행하였다. 레이블링을 통해 표 1과 같은 규모의 데이터세트가 구축되었으며, 구축된 데이터세트는 이미지와 레이블이 쌍을 이룬다. 레이블에는 이미지 내 포함된 객체들의 클래스 식별자와 0-1 사이의 실수로 정규화된 바운딩 박스 정보가 기록되어 있다.

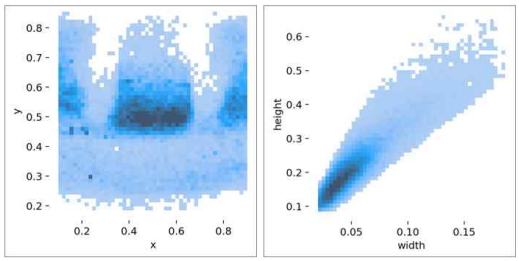

그림 1. 사람 인스턴스에 대한 바운딩 박스의 분포(좌) 및 크기(우)
Fig. 1. Distribution(left) and size(right) of the bounding boxes for the person instances

구축된 데이터세트에서 사람으로 레이블링된 인스턴스들은 그림 1과 같은 분포와 크기를 보였다. 사람 인스턴스들은 이미지 내에서 정중앙과 좌우에 주로 포착되었으며, 그 외의 영역에서는 고른 분포로 나타났다. 다만 이미지 내에서 위쪽으로는 거의 사람이 나타나지 않았는데, 이는 지게차의 마스트(Mast)로 인해 이미지를 수집하는 카메라 시야가 가려져있기 때문이다. 다음으로 이미지 내 사람의 바운딩 박스 크기의 경우 너비는 0.07 ± 0.027, 높이는 0.25 ± 0.084를 보였으며 바운딩 박스의 비율(너비/높이)은 0.49 ± 0.07로 나타났다.

3.2 GAN 기반 사람 이미지 생성

본 논문에서는 가상의 사람 이미지를 생성하기 위하여 StyleGAN2[10] 아키텍처를 기반으로 그림 2와 같은 생성자 네트워크를 설계하였다.

생성자 네트워크는 일반적인 GAN과 다르게, 잠재 벡터 z로부터 바로 특징(feature)을 생성하지 않고 매핑 네트워크(Mapping Network) f를 통해 잠재 벡터 z를 W 공간에 매핑한 뒤에 이를 사용해 특징을 생성하는 구조로 되어있다. 일반적으로 고정된 분포(보통 정규분포)로부터 샘플링된 z는 특정 값을 변화시키면 매우 급격한 특징 변화를 보이는 반면 f(z)는 고정된 분포를 따르지 않으므로 w의 일부 값이 변화하여도 비교적 선형적인(linear) 특징 변화가 나타난다. 이처럼 잠재 공간이 표현(representation)에 따라 정렬되는 현상을 Distentanglement[10]라고 불리우며, 이러한 선형적인 특징 변화는 데이터 증강과 같이 특징을 제어하여 원하는(부족한) 데이터 생성을 수월하게 하며, 동시에 급격한 특징 변화로부터 유발되는 이미지 충실도의 변화를 완화시키는데 도움을 줄 수 있다.

본 논문에서는 MLP(Multi-Layer Perceptron)로 구성된 SytleGAN2의 매핑 네트워크를 포지셔널 임베딩(Positional Embedding)과 닷-프로덕트 어텐션(Dot-product Attention)으로 구성된 트랜스포머 구조로 변경하였다. 일반적으로 트랜스포머 구조는 음성, 텍스트와 같이 순서가 있는 데이터에서 맥락을 학습하는데 좋은 성능을 보이므로 자연어 처리 분야에서 널리 사용되고 있다. 그리고 최근에는 이미지 인식 분야에서도 이미지를 여러 패치로 나누고 순차적으로 나열하여 학습하는 트랜스포머 구조의 분류기[21]가 등장하면서 비전 분야 전반에서도 각광 받고있다. 본 논문에서는 앞서 언급한 잠재 공간이 표현에 따라 Disentanglement되도록 유도하기 위하여 트랜스포머 구조의 f을 제안한다. f의 구조는 다음과 같다. 먼저 정규 분포로부터 추출된 512 차원의 잠재 벡터 z를 MLP에 통과시켜 4,096 차원의 특징을 도출하고, 256x16 크기의 행렬로 변환(reshape)한다. 이는 16차원의 벡터 256개가 나열된 형태로 볼 수 있으며, 포지셔널 임베딩을 사용해 일련의 순서를 부여한다. 그리고 닷-프로덕트 어텐션 레이어를 통과시켜 표현에 대한 가중치를 부여한다. 이렇게 산출된 특징은 포지셔널 임베딩의 출력과 원소 합(element-wise Sum)을 계산하여 MLP에 입력하며, 다시 한 번 포지셔널

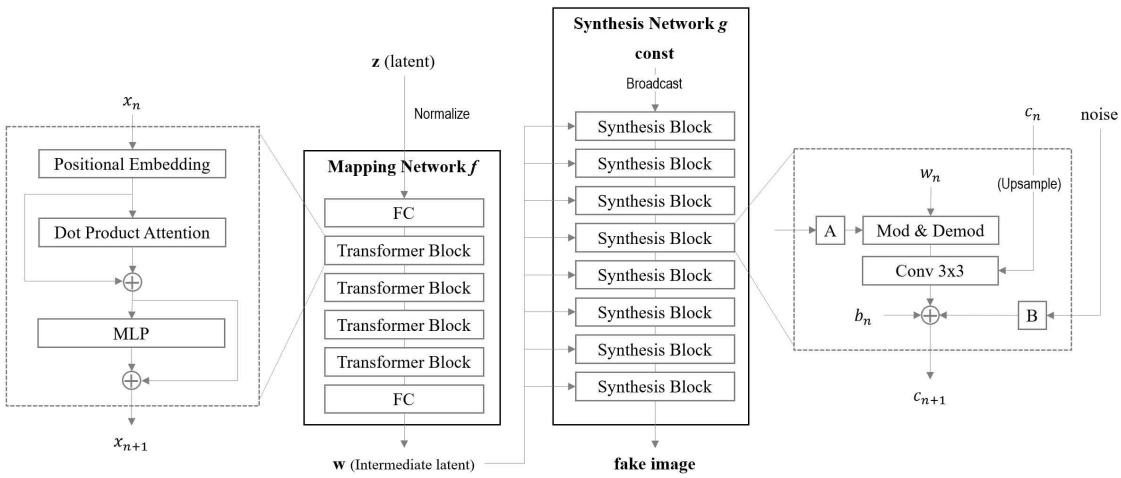

그림 2. 본 논문에서 제안하는 생성자 네트워크의 구조
Fig. 2. The overall structure of the proposed generator network

임베딩 레이어의 출력과 원소 합을 계산하여 MLP를 통해 순서 정보가 사라지는 것을 방지한다. 마지막으로 매핑 네트워크 f는 앞서 설명한 트랜스포머 블록을 4개를 쌓은 구조이며, 매핑 네트워트의 마지막 MLP는 512 차원의 출력을 가지도록 설정하여 잠재 벡터 z와 동일한 쉐입을 가진다. 매핑 네트워크를 통해 도출된 매개 잠재 벡터 w는 z와는 다르게 보다 선형적인 분포를 가질 것이며, 8x64 크기의 행렬로 변환하여 각 행을 합성 블록(Synthesis Block)에 입력함으로써 각 행의 벡터들이 저차원부터 고차원까지의 특징을 생성하는데 영향을 미치도록 유도한다. 즉, w의 각 열의 변화는 사람의 손, 발, 머리카락 등의 작은 특징 변화를 유도하고, 각 행의 변화는 사람의 몸집, 포즈 등의 큰 특징 변화와 관련이 높아진다. 그 외 판별자 네트워크는 기존 StyleGAN2와 동일하게 구성하였다.

일반으로 GAN은 주어진 이미지 내에 특정 객체에 집중하여 이미지를 생성하지 않고, 주어진 이미지와 전체적으로 유사한 장면(scene)를 생성한다. 따라서 비슷한 위치에 비슷한 크기로 존재하는 특징은 잘 학습하지만, 임의 위치에 다양한 크기로 존재하는 특징은 비교적 학습이 어렵다. 특히 본 연구에서 다루는 산업현장의 경우, 여러 경로에서 사람이 등장하기 때문에 이미지 내에 사람의 크기가 매우 다양하다. 표 1의 데이터세트에서 사람의 바운딩 박스를 픽셀 단위로 환산하면 너비는 90.12±34.81, 높이는 181.26±60.85로 너비와 높이 모두 편차가 크다. 게다가 그림 1과 같이 이미지 내에 사람이 놓인 위치도 다양하다. 따라서 본 논문에서 다루는 산업현장의 데이터와 같이, GAN을 통해 이미지 내에 위치 및 크기의 편차가 큰 동일 객체를 학습하기 위해서는 GAN이 학습하는 입력 공간(input space)의 크기를 제한할 필요가 있다. 본 논문에서는 바운딩 박스에 해당하는 이미지를 잘라내고(copping), 0으로 채워진(zero-padded) 빈 이미지 정중앙에 배치하여 GAN 학습을 위한 사람 이미지를 만들었다. 따라서 생성자 네트워크는 512x512x3의 입력 공간의 중심을 기준으로 학습이 이루어질 것이며, 커널을 위에서 아래로 이동하며 이미지를 생성하는 CNN 구조의 생성자 네트워크에게 엣지 이펙트(edge-effect) 유도하여 0으로 채워진 배경과 생성하는 이미지의 경계를 자연스럽고 뚜렷하게 생성할 수 있을 것이다. 본 논문에서는 표 1의 데이터세트에서 Train으로 구분된 78,505개의 사람 인스턴스를 대상으로, 앞서 설명한 방법에 따라 그림 3과 같은 512x512 해상도의 컬러 이미지 78,505장을 만들어 GAN의 학습 데이터로 사용하였다.

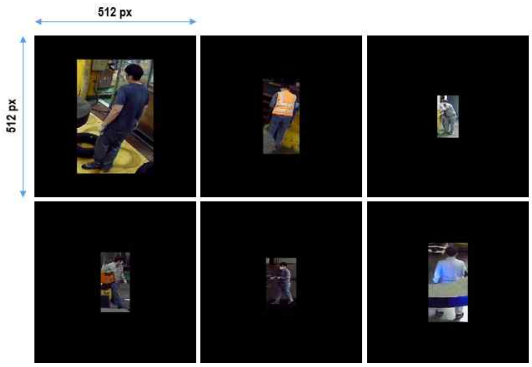

그림 3. GAN 학습을 위한 사람 이미지 샘플
Fig. 3. Person image samples for GAN training

StyleGAN2와 본 논문에서 제안하는 GAN 모두 아래와 같은 환경에서 학습을 진행하였으며, 32의 배치 크기로 78,505장의 이미지를 180번 반복하여 학습(160 epochs)하기까지 두 모델 모두 약 5일의 시간이 소요되었다.

- OS: Ubuntu 20.04 LTS
- CPU: 2 x 12-core Intel Gold 5118
- GPU: 8 x NVIDIA TITAN Xp
- Deep Learning Framework: Flax 0.5.3

GAN을 통해 생성된 데이터가 실제 데이터와 얼마나 유사한지 평가하기 위해 GAN 연구자들은 다양한 평가 지표를 제안하였으며, 지표에 따라 장단점이 존재한다[22]. 본 논문에서는 GAN 기반 이미지 생성 분야에서 생성된 이미지의 충실도를 측정하기 위한 지표로 널리 사용되고 있는 FID(Fréchet Inception Distance)[23]를 사용하여, 본 논문에서 제안하는 GAN과 StyleGAN2가 생성한 이미지의 충실도를 비교한다. FID는 수식 (1)과 같이 계산한다. r과 g는 각각 실제 이미지(GAN 학습에 사용하지 않은)와 생성된 이미지를 Inception-v3[24]에 입력하여 나온 마지막 평균 풀링(Average Polling) 레이어의 2,048 차원의 특징을 의미한다. (μ_r, Σ_r)과 (μ_g, Σ_g)는 r과 g에 대한 평균과 공분산이며, Tr은 대각 성분의 합을 의미한다. 본 논문에서는 GAN 학습에 사용하지 않은 표 1의 Test를 사용하여 제안하는 GAN과 StyleGAN2가 생성한 이미지의 충실도를 평가하였다. FID는 두 그룹의 특징 차이를 계산하는 것으로 GAN이 생성한 이미지가 실제 이미지와 유사할수록 0에 가까운 점수를 보인다.

$$FID(r,g) = \|\mu_r - \mu_g\|_2^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}}) \quad (1)$$

그림 4-(위)는 제안하는 GAN과 StyleGAN2의 학습 진행에 따라 생성된 이미지의 FID 변화를 나타낸 그래프이다. 두 모델 모두 1 에폭(epoch) 단위로 임의의 이미지를 10,000장을 생성하고, 학습에 사용하지 않은 표 1의 Test 10,115장을 사용하여 수식 (1)에 따라 FID를 측정하여 기록하였다. 두 모델 모두 20 에폭까지는 통계적으로 유의미한 FID 차이를 보이지 않았으나, 이후에는 근소하게 제안하는 GAN의 FID가 낮아지는 것을 확인하였다. 제안하는 GAN의 최저 FID는 18.52를 기록하였으며, StyleGAN2는 31.05의 FID를 기록하였다.

그림 4-(아래)는 제안하는 GAN이 생성한 이미지에 대하여, 측정된 FID 차이에 따른 충실도를 비교한 결과이다. 각각의 생성된 이미지는 동일한 잠재 벡터로부터 생성된 가상의 사람 이미지이며, FID가 낮아질수록 인간의 시각으로 바라볼 때 사람이라고 인식될 수 있는 얼굴, 팔, 다리 등의 특징이 점점 명확해지는 경향을 확인할 수 있다.

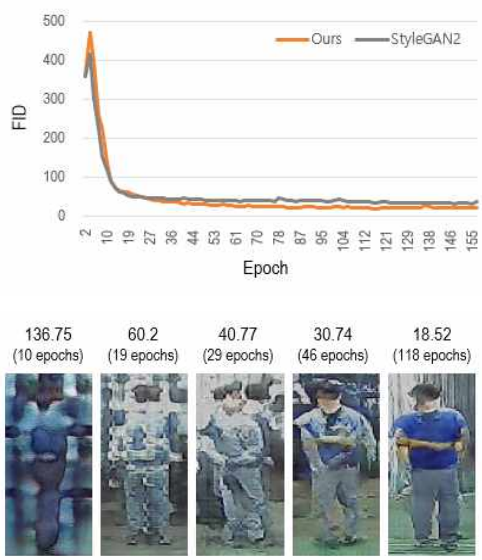

그림 4. (위) 학습 진행에 따른 FID 점수 변화 (아래) FID 점수 차이에 따른 생성된 이미지의 충실도 비교

Fig. 4. (Top) FID score changes with training progresses (Down) Comparison of fidelity of generated images according to FID score differences

3.3 GAN 기반 사람 인스턴스 증강

본 절에서는 제안하는 GAN을 통해 생성한 가상의 사람 이미지를 사용하여 기존 데이터세트를 증강하는 방법을 다루며, 전반적인 개념도는 그림 5와 같다. 본 논문에서 제안하는 GAN은 정중앙에 가상의 사람 인스턴스가 위치하고 나머지는

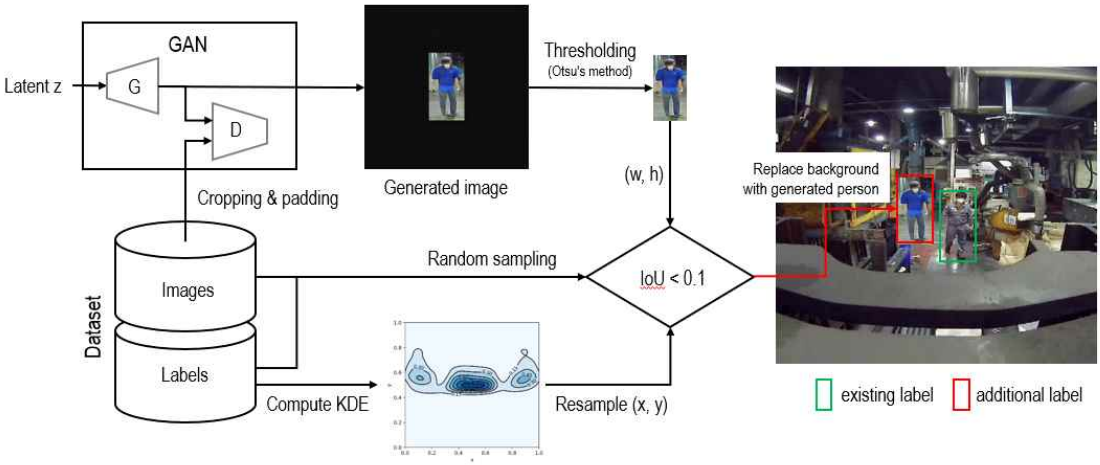

그림 5. 제안하는 GAN 기반의 이미지 데이터 증강 개념도
Fig. 5. An illustration of our proposed GAN-based image data augmentation

0에 가까운 값으로 채워진 512x512 해상도의 컬러 이미지를 생성한다. 본 논문에서는 생성된 이미지로부터 가상의 사람 인스턴스를 분리하기 위해 이미지 전체의 픽셀 분포를 바탕으로 임계값을 찾아 분리하는 오츠 알고리즘[25]을 사용하였다. 분리된 사람 인스턴스는 너비와 높이(w, h)를 산출하여 바운딩 박스의 정보로 사용한다. 다음으로 분리된 가상의 사람 인스턴스가 실제 데이터세트의 이미지 내에 어느 곳에 배치될지 결정하기 위하여, 실제 데이터세트에 기록된 사람 객체의 바운딩 박스 중심 좌표들에 대하여 KDE (Kernel Density Estimation)을 수행하고 확률 밀도 함수를 도출한다. 도출된 확률 밀도함수를 사용해 임의의 좌표 (x, y)를 다시 샘플링 함으로써 앞서 설명한 가상의 사람 인스턴스에 대한 바운딩 박스 정보(x, y, w, h)를 자동으로 생성할 수 있다. 마지막으로 기존 데이터세트에 가상의 사람 인스턴스를 삽입하기 위하여, 임의로 추출된 이미지-레이블 쌍에 대하여 기존 바운딩 박스와 가상의 사람 인스턴스의 바운딩 박스가 겹치지 않는지 확인한다. 본 논문에서는 두 바운딩 박스의 IoU(Intersection over Union)가 0.1 이하인 경우에만 가상의 사람 인스턴스를 기존 데이터세트에 추가하도록 설정하였다.

4. 증강된 데이터의 객체 검출 영향력 평가

4.1 베이스라인

3장을 통해 제안한 GAN 기반 이미지 데이터 증강이 실제 딥러닝 객체 검출 모델에 미치는 영향력을 평가하기 위하여, YOLOv5를 본 영향력 평가의 베이스라인으로 설정하였다. YOLOv5는 학습 과정에 스케일링(scaling)과 같은 기존 이미지 조작 기법들을 조합하여 주어진 학습 데이터를 증강하는 기법이 포함되어 있다. 특히 4개의 학습 데이터를 임의로 선택하고 하나의 이미지처럼 연결하는 모자이크(Mosaic) 증강의 경우, 인스턴스의 다양성과 절대적인 개수를 크게 증가키는 효과가 있으며, 주어진 학습 데이터에 클래스 불균형 문제가 있더라도 과적합 문제로부터 비교적 자유로운 장점이 있다. YOLOv5는 클래스 불

표 2. 생성된 사람 인스턴스 추가에 따른 객체 검출 성능 비교
Table 2. Comparison of object detection performance by adding generated person instances

	# generated instances (person)	Precision	Recall	mAP50	mAP50-90	TP (person)	FN (background)
Baseline	-	0.923	0.916	0.953	0.709	0.8016	0.0508
w/ Augmentation	7,500	0.924	0.913	0.951	0.709	0.7994	0.0522
	15,000	0.919	0.914	0.953	0.709	**0.8106**	**0.0453**
	22,500	0.92	0.914	0.95	0.705	0.8041	0.0517
	30,000	0.92	0.917	0.953	0.708	0.8035	0.0469

균형이 있는 표 1의 데이터세트를 학습하였음에도 mAP50 0.95 이상을 기록하였다. YOLOv5는 전통적인 이미지 조작 기반의 데이터 증강만으로 충분한 성능을 보였으므로, 본 논문에서 제안하는 GAN 기반 이미지 데이터 증강이 추가적으로 성능 개선에 기여할 수 있는지 평가하기 적절한 베이스라인으로 볼 수 있다.

본 논문에서는 MS COCO[26] 데이터세트를 사전학습한(pretrained) YOLOv5s 모델에 표 1의 Train을 추가 학습(fine-tuning) 시키고 가장 높은 mAP를 기록한 모델을 베이스라인으로 설정하였다. 추가 학습은 배치 크기를 256으로 설정하고 4개의 NVIDIA TITAN Xp를 사용하여 약 2일의 시간이 소요되었다. 그리고 표 1의 Test에 대하여 평가한 검출 성능은 6개 클래스에 대하여 mAP50 0.953, mAP50-95 0.709, 사람에 대한 TP(True Positive) 0.8016을 보였다. 이와 같은 베이스라인의 객체 검출 성능은 본 논문의 영향력을 가늠하기 위한 기준점으로 사용한다.

4.2 검출 성능에 미치는 영향력 평가

본 절에서는 제안하는 GAN을 통해 생성된 가상의 사람 인스턴스를 객체 검출 모델의 학습 데이터에 추가함에 따라 검출 성능이 얼마나 개선되는지를 평가한다. 평가 방법은 다음과 같다. 본 논문에서 제안한 GAN에 임의의 잠재 벡터 n개를 입력하여 가상의 사람 인스턴스 n개를 생성하고, 3.3절의 데이터 증강 방법을 사용하여 기존 데이터세트에 가상의 사람 인스턴스를 추가한다. 사람 인스턴스가 증강된 데이터세트는 베이스라인과 동일한 YOLOv5s를 사용해 200 에폭 이내로 학습을 진행하고, 표 1의 Validation에 대하여 최고의 mAP를 보이는 모델을 도출하며, 표 1의 Test에 대한 검출 성능을 비교한다. 검출 성능의 mAP 측정을 위한 IoU 임계값은 0.6으로 설정하였으며, 혼동 행렬(confusion matrix)를 위한 신뢰도(confidence) 임계값은 0.25로 설정하였다. n을 7,500(Train의 약 10%) 단위로 증가시켰을 때 나타난 검출 성능의 변화는 표 2와 같이 측정되었다.

먼저 본 논문에서 제안하는 GAN 기반의 데이터 증강을 수행하여도 mAP 측면에서 유의미한 성능 하락은 발견되지 않았다. 즉, 기존 데이터세트에 가상의 사람 인스턴스를 추가하여 학습하여도 객체 검출 모델이 어떠한 객체의 위치를 특정(localization)하는 성능에는 영향을 미치지 않는다고 해석할 수 있다. 그러나 바운딩 박스의 객체가 무엇인지를 분류(classification)하는 성능에는 영향을 주었다. 표 2의 사람에 대한 TP와 같이 n이 15,000(Train의 약 20%)일 때 사람을 사람으로 더 잘 예측하였다. 이러한 성능 개선 효과가 나타난 이유는 그림 6의 혼동 행렬을 통해

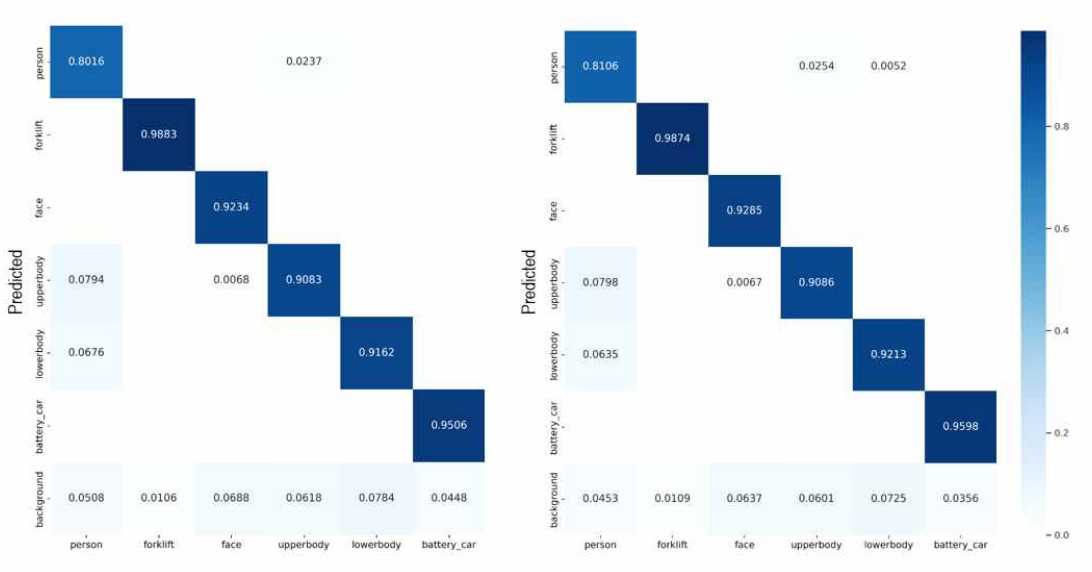

그림 6. (좌) 베이스라인 모델의 혼동 행렬 (우) 생성된 사람 인스턴스 15,000개가 포함된 증강된 데이터를 학습한 모델의 혼동 행렬

Fig. 6. (Left) Confusion matrix of the baseline model (Right) Confusion matrix of the model that trained augmented data containing 15,000 generated person instances

설명할 수 있다. FN(False Negative)과 같이 실제 정답은 사람이지만 신뢰도 임계값 미만으로 예측하지 못한(background) 비율과 실제 사람이지만 하체(lower body)라고 잘못 예측한 비율이 감소했기 때문이다.

마지막으로 n에 따라 오히려 객체 검출 모델의 성능이 낮아질 수 있다는 결과도 주목할 필요가 있다. 만약 GAN이 생성한 이미지의 충실도가 실제 데이터와 같은 수준이라면 n과 검출 성능은 비례할 것이다. 하지만 현실적으로 GAN을 통해 생성된 이미지에는 실세계에서 관측할 수 없는 이미지도 포함될 수 있으며(제한된 데이터세트로부터 실제 분포를 추정한 것이므로), n에 비례하여 그 수가 많아진다면 오히려 객체 검출 모델의 분류 성능은 떨어질 수 있으므로 적절한 수준의 n을 찾는 작업이 요구된다.

5. 결 론

본 논문은 산업현장에서 매우 중요한 객체인 사람에 초점을 맞추어 GAN 기반 가상의 사람 이미지를 생성하고 이를 활용해 딥러닝 객체 검출 모델의 성능을 개선하는 데이터 증강을 제안하였다. 본 논문에서 제안한 생성자 네트워크는 기존 방법론보다 높은 충실도의 이미지를 생성하였으며, 생성된 이미지로부터 객체를 분리하고 자동으로 바운딩 박스 정보를 부여하여 기존 데이터세트에 가상의 객체 인스턴스를 추가하는 데이터 증강을 제시하였다. 그리고 제안하는 데이터 증강이 실제 딥러닝 객체 검출 모델에 미치는 영향력을 평가하였으며, 객체 검출 모델의 수정 없이 본 논문에서 제안하는 증강만을 사용하여 성능을 개선할 수 있음을 보였다. 본 논문의 결

과는 사람과 각종 중장비가 혼재된 건설 및 제조 현장의 산업재해를 예방하는 인공지능 기술에 활용될 수 있을 것이다.

향후 연구에서는 보다 높은 충실도의 이미지 생성이 가능하도록 생성자 네트워크를 개선할 것이며, 이미지의 분할(segmentation) 정보를 사용하여 객체와 배경을 분리하여 학습하고 자연스럽게 배경과 객체를 합성하는 GAN에 관한 연구를 수행할 계획이다.

본 연구는 산업통상자원부와 한국산업기술진흥원의 “국가혁신클러스터사업(P0015279_산업현장 내 맞춤형 AI 서비스를 위한 지능형 플랫폼 개발 및 실증 운영)”의 지원을 받아 수행된 연구결과임

참 고 문 헌

[1] Benyang D., Xiaochun L., and Miao Y., “Safety helmet detection method based on YOLO v4”, 16th International Conference on Computational Intelligence and Security, pp. 155-158, 2020.
DOI: 10.1109/CIS52066.2020.00041

[2] Nath Nipun D., Amir H. Behzadan, and Stephanie G. Paal., "Deep learning for site safety: Real-time detection of personal protective equipment", Automation in Construction, vol. 112, 2020.
https://doi.org/10.1016/j.autcon.2020.103085

[3] Zhao, congcong, and Chen B., "Real-time pedestrian detection based on improved YOLO model", 11th International Conference on Intelligent Human-Machine Systems and Cybernetics, vol. 2, 2019..
DOI: 10.1109/IHMSC.2019.10101

[4] Xu H., Guo M., Nedjah N., Zhang J., and Li P., “Vehicle and pedestrian detection algorithm based on lightweight YOLOv3-promote and semi-precision acceleration”, IEEE Transactions on Intelligent Transportation Systems, vol. 23, pp. 19760-19771, 2022.
DOI: 10.1109/TITS.2021.3137253

[5] Halevy, Alon, Peter N., and Fernando P., "The unreasonable effectiveness of data", IEEE intelligent systems, vol. 24, pp. 8-12, 2009. DOI: 10.1109/MIS.2009.36

[6] Sun C., Shrivastava A., Singh S., and Gupta A., "Revisiting unreasonable effectiveness of data in deep learning era“, Proceedings of the IEEE international conference on computer vision. 2017.
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8237359

[7] Ahmed M., Hashmi K. A., Pagani A., Liwicki M., Stricker D., and Afzal, M. Z., "Survey and performance analysis of deep learning based object detection in challenging environments", Sensors, vol. 21, 2021. https://doi.org/10.3390/s21155116

[8] ImageNet Large Scale Visual Recognition Challenge (ILSVRC),
http://www.image-net.org/challenges/LSVRC/

[9] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., ... and Bengio Y., “Generative adversarial networks”, Communications of the ACM, vol. 63, pp.139-144, 2020.
https://doi.org/10.1145/3422622

[10] Karras T., Laine S., Aittala M., Hellsten J., Lehtinen J., and Aila, T., “Analyzing and improving the image quality of stylegan", Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8110-8119, 2020.
https://arxiv.org/abs/1912.04958

[11] Zhao L., Zhang Z., Chen T., Metaxas D., and Zhang H., "Improved transformer for high-resolution gans", Advances in Neural Information Processing Systems, vol. 34, pp. 18367-18380, 2021.

https://arxiv.org/abs/2106.07631
[12] Frid-Adar M., Diamant I., Klang E., Amitai M., Goldberger J., and Greenspan H., "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification", Neurocomputing, vol. 321, pp. 321-331, 2018. https://doi.org/10.1016/j.neucom.2018.09.013
[13] Salehinejad H., Valaee S., Dowdell T., Colak E., and Barfett J., "Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks", 2018 IEEE international conference on acoustics, speech and signal processing, pp. 90-994, 2018. DOI: 10.1109/ICASSP.2018.8461430
[14] Wang C., and Xiao Z., "Lychee surface defect detection based on deep convolutional neural networks with GAN-based data augmentation", Agronomy, vol. 11, 2021. https://doi.org/10.3390/agronomy11081500
[15] Jiang Y., Chang S., and Wang Z. “Transgan: Two transformers can make one strong gan”, arXiv preprint arXiv:2102.07074, 2021. https://doi.org/10.48550/arXiv.2102.07074
[16] Zhao S., Liu Z., Lin J., Zhu J. Y., and Han S., "Differentiable augmentation for data-efficient gan training", Advances in Neural Information Processing Systems, vol. 33, pp. 7559-7570, 2020. https://doi.org/10.48550/arXiv.2006.10738
[17] Liu B., Su S., and Wei J., "The effect of data augmentation methods on pedestrian object detection", Electronics, vol. 11, 2022. https://doi.org/10.3390/electronics11193185
[18] Inoue H., "Data augmentation by pairing samples for images classification", arXiv preprint arXiv:1801.02929, 2018. https://doi.org/10.48550/arXiv.1801.02929
[19] He K., Zhang X., Ren S., and Sun J., "Deep residual learning for image recognition", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016. https://doi.org/10.48550/arXiv.1512.03385
[20] YOLOv5(Ultralytics) https://github.com/ultralyti cs/yolov5
[21] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., and Houlsby N., "An image is worth 16x16 words: Transformers for image recognition at scale" arXiv preprint arXiv:2010.11929, 2020. https://doi.org/10.48550/arXiv.2010.11929
[22] Borji A., "Pros and cons of gan evaluation measures", Computer Vision and Image Understanding, vol. 179, pp. 41-65, 2019. https://doi.org/10.1016/j.cviu.2018.10.009
[23] Heusel M., Ramsauer H., Unterthiner T., Nessler B., and Hochreiter S., "Gans trained by a two time-scale update rule converge to a local nash equilibrium", Advances in neural information processing systems, vol. 30, 2017. https://doi.org/10.48550/arXiv.1706.08500
[24] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., and Wojna, Z., "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826, 2016. DOI: 10.1109/CVPR.2016.308
[25] Otsu N., "A threshold selection method from gray-level histograms", IEEE transactions on systems, man, and cybernetics, vol. 9, pp.62-66, 1979. DOI: 10.1109/TSMC.1979.4310076
[26] Lin T. Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., ... and Zitnick C. L., "Microsoft coco: Common objects in context", European conference on computer vision. Springer, Cham, pp. 740-755, 2014. https://doi.org/10.48550/arXiv.1405.0312

저 자 소 개

백문기(Moon-Ki Back)

2013.2 충남대학교 컴퓨터공학과 졸업
2015.2 충남대학교 컴퓨터공학과 석사
2021.2 충남대학교 컴퓨터공학과 박사
2021.3-현재 한국전자통신연구원 박사후연구원
<주관심분야> 딥러닝, 생성 모델, 데이터 증강, 객체 검출, 음성 인식

김계경(Kye-Kyung Kim)

1989.2 경북대학교 전자공학과 졸업
1992.2 경북대학교 전자공학과 석사
1997.2 경북대학교 전자공학과 박사
1998.8-2001.2 CENPARMI 방문과학자
2001.3-현재 한국전자통신연구원 근무
<주관심분야> 패턴인식, 로봇비전, 컴퓨터 비전, 인공지능