

# 객체 탐지 성능 향상을 위한 생성형 인공지능 기반 데이터 증강 기법 연구

김도희<sup>○</sup>, 김명호<sup>\*</sup>

<sup>○</sup>숭실대학교 소프트웨어학과,

<sup>\*</sup>숭실대학교 소프트웨어학과

e-mail: ksmd09089797@gmail.com<sup>○</sup>, kmh@ssu.ac.kr<sup>\*</sup>

## A Study on Generative Artificial Intelligence-Based Data Augmentation Techniques for Enhancing Object Detection Performance

Dohee Kim<sup>○</sup>, Myongho Kim<sup>\*</sup>

<sup>○</sup>Dept. of Software, Soongsil University,

<sup>\*</sup>Dept. of Software, Soongsil University

### ● 요약 ●

최근 딥러닝 기술의 발달로 물체 탐지를 위한 객체 인식 분야가 기계학습을 접목한 연구가 급격히 증가하고 있다. 하지만, 탐지하려는 물체가 다른 객체에 가려진 경우와 같이 특수한 상황에 대한 데이터의 수량이 부족하여 성능 저하를 야기한다는 점과, 객체 탐지 수행 과정에서 작은 객체의 탐지가 어렵다는 한계점이 있다. 본 연구는 진술한 문제점을 보완할 방법을 제안한다. 데이터 증강 기법을 이용하여 클래스가 부족한 데이터의 양을 늘려 학습 데이터를 증강시켰다. 한편, SRGAN을 사용하여 작은 객체를 확대시킨 뒤 이미지를 합성시켜 데이터를 구성하였다. 제안된 방법은 PyTorch 환경에서 YOLOv5를 수행한 결과, 객체 탐지 성능이 향상되는 것을 확인할 수 있었다.

**키워드:** 딥러닝(Deep Learning), 객체 탐지(Object Detection), 데이터 증강(Data Augmentation), GAN

## I. Introduction

객체 탐지(Object Detection)는 컴퓨터 비전 분야의 핵심 연구 주제로, 이미지 또는 비디오에서 특정 객체의 위치와 종류를 분류하는 기술이다. 딥러닝(Deep learning) 기술의 발전으로 인해 객체 탐지 기술은 여러 분야에서 활용되고 있다[1]. 하지만 객체 탐지 알고리즘은 화면에 전부 담기지 않아 객체의 일부만 촬영되거나, 물체에 가려지는 경우 인식이 떨어진다 단점이 있다. 딥러닝 모델은 학습 데이터에 따라 성능이 좌우되므로 양질의 데이터셋을 사용하는 것이 중요하지만, 이 같은 특정상황에 대한 학습 데이터셋이 부족하기 때문이다. 뿐만 아니라 딥러닝 기반의 객체 탐지 알고리즘의 고전적인 문제점으로, 객체 탐지 수행 중 업샘플링(Upsampling) 과정으로 국소적인 정보가 사라져 작은 객체를 잘 예측하지 못한다는 한계점이 있다[2].

따라서 본 논문은 객체 탐지 향상을 위해 딥러닝 기반 생성형 인공지능을 활용한 데이터 증강 기법 연구를 제안한다. 생성형 인공지능의 일종인 GAN(Generative Adversarial Networks)[3]을 통해

클래스가 부족한 데이터의 양을 늘려 학습 데이터셋을 증강시킨다. 또한, 이미지 데이터의 그리드(Grid)를 분할하여 고해상도 텍스처(Texture)로 변환한 뒤, 분할된 이미지를 합성하여 객체 탐지 알고리즘의 작은 객체 탐지 성능을 향상시킨다.

본 논문의 구성은 다음과 같다. 2장에서 딥러닝 기반의 객체 탐지 기술과 이미지에 대한 고전적인 데이터 증강 방법과 생성형 AI의 기존 연구에 대해 알아본다. 3장에서 객체 탐지 성능 향상을 위한 생성형 인공지능 기반 데이터 증강 기법에 대한 제안 구조 및 연구계획을 소개한다. 4장에서 3장에서 증강된 데이터를 이용한 객체 탐지 실험 결과를 보여주고 5장에서 결론을 맺는다.

## II. Related works

### 1. Object Detection

#### 1.1 YOLOv5

YOLO(You Only Look Once)[4]계열 알고리즘은 빠르고 정확한 객체 탐지 알고리즘이다. 기존의 탐지 방법들이 먼저 후보 영역을 선택하고, 그 영역을 분류하는 두 단계 과정을 거치는 반면, YOLO는 이미지를 한 번만 보고 객체의 위치와 클래스를 동시에 예측한다. YOLO계열 알고리즘은 이미지를 그리드로 나누고, 각 그리드 셀(cell)이 하나 이상의 객체를 감지하도록 한다. 각 셀은 여러 개의 경계 상자(Bounding Boxes)를 예측하며, 이 상자들은 각각 객체의 위치와 크기, 클래스에 대한 확률을 예측한다. 하지만 YOLO계열 알고리즘은 작은 객체에 대한 예측 정확도가 낮다는 단점이 있다.

### 2. Data Augmentation

#### 2.1 Classical Method

데이터 증강 기법은 새로운 데이터를 수집하지 않고 기존의 데이터를 기하학적 변형을 통해 데이터의 수를 늘리는 방법이다. 이미지 데이터의 경우 색상 변환, 회전, 왜곡, 확대 및 축소, 분할 등이 있다. 본 논문에서는 이미지의 분할과 확대, GAN을 활용하여 데이터를 증강시켰다.

#### 2.2 DCGAN

기존의 GAN에 CNN을 결합한 DCGAN(Deep Convolutional Generative Adversarial Network)[5]은 생성기(generator)와 판별기(discriminator)로 구성된다. 생성기는 랜덤 노이즈(Random noise)를 받아 실제와 유사한 이미지로 변환하는 역할을 한다. 판별기는 실제 이미지와 생성기가 만든 가짜 이미지를 받아서 그것이 실제 이미지인지 가짜 이미지인지를 판별하는 역할을 한다. 이 같은 구조를 가진 DCGAN은 이미지를 생성하는데 뛰어난 성능을 보인다.

#### 2.3 Super Resolution - SRGAN

SRGAN(Super-Resolution Generative Adversarial Network)[6]는 저해상도 이미지를 고해상도 이미지로 변환하는 과정에서 텍스처 디테일(Texture details)에 관한 손실 없이 광학적 실재성을 보존하는 초고해상도 이미지 생성 방법이다. 이 같은 SRGAN은 저해상도 이미지를 고해상도 이미지로 업스케일(Upscale)하는 생성기와 생성된 고해상도 이미지가 실제 이미지와 가짜 이미지를 구분하는 판별기로 이루어졌다. 판별 과정을 통해 생성기는 판별기를 속이는 방향으로 점차 학습하게 되고, 이런 경쟁적인 학습 과정을 통해 SRGAN은 저해상도 이미지를 현실적인 고해상도 이미지로 변환하는 능력을 향상시킨다.

## III. The Proposed Scheme

본 논문은 객체 탐지 수행 시, 특정 상황에 대한 학습 데이터셋의 클래스가 부족하다는 문제점과 작은 객체의 탐지가 어렵다는 한계점을 극복하기 위해 GAN을 활용한 데이터셋 증강 기법을 제안한다. 이미지 데이터에서 추출한 분할된 그리드는 Super Resolution 기법을 적용하여 고해상도 텍스처로 변경 후, 합성하여 신경망에 입력값으로 적용한다.

본 연구는 이미지 분석에 사용되는 대규모 데이터셋인 COCO(Common Objects in Context) [7] 데이터셋을 축소하여 사용한다. 80개의 카테고리를 가진 200,000개 이상의 레이블링(Labeling)된 이미지로 구성되며, 하나의 이미지에 크고 작은 하나 이상의 객체가 포함되어있다. 가려지거나 작게 나온 사람의 탐지 성능을 향상시키는 본 연구 목적에 맞게 COCO 데이터셋에서 500개의 이미지를 추출하여 사람 외 다른 클래스를 없애고 사람에 대한 레이블링 작업을 수행하여 학습 데이터셋을 구성한다. 학습 데이터셋은 크고 작은 크기로 표현된 사람, 온전한 형태의 사람, 물체에 가려져 상반신 또는 하반신만 나온 사람, 팔 또는 다리 등 일부 신체만 나온 사람을 포함한다.

위 과정을 통해 생성된 데이터는 이전 과정에서 학습이 완료된 YOLOv5 모델에 입력값으로 사용하여 객체 탐지를 수행한다.

학습 데이터셋은 이미지 분할, 확대, DCGAN을 통해 1,000개로 증강시켜 YOLOv5 모델에 학습시킨다. 학습된 모델의 성능평가는 정밀도(Precision)와 재현율(Recall)을 측정하여 mAP(Mean Average Precision)을 산출한다. 다음 식 1은 정밀도와 재현율을 계산하는 수식이다.

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \quad \dots (1)$$

Precision은 모델이 True라고 예측한 것 중 실제로 True인 것의 비율을 나타내며, Recall은 실제 True인 것 중 모델이 True라고 예측한 것의 비율을 나타낸다. mAP는 각 클래스의 AP(Average Precision)를 계산한 후, 클래스에 대한 평균을 계산한다. AP는 Precision-Recall 그래프의 밑면적의 넓이다.

한편, COCO 데이터셋의 이미지를 SRGAN를 통해 고해상도 텍스처로 변경한 Super Resolution 데이터셋을 생성한다. 제안 기법은 입력 이미지에 대하여 일정 크기의 그리드 셀로 분할한다. 그 뒤 그리드 셀을  $N^2$ 크기로 확대하고, 확대한 각 셀에 대한 합성을 수행하여 원본 이미지와 똑같은 특성을 가진 데이터를 생성한다. 본 연구 수행 환경은 다음 표 1과 같다.

Table 1. Experiment Environment

OS	Window 10 Pro
CPU	AMD Ryzen 5 3600
RAM	32 GBytes
Graphic card	NVIDIA GeForce RTX 2070 SUPER
Program	CUDA 11.7
Library	PyTorch 2.0.1

Table 2. Experiment Result

epoch: 100 (%)	Precision	Recall	mAP_0.5	mAP_0.5 :0.95
Data 500	93.7	78.9	87.58	63.79
Data 1000	93.34	80.76	88.58	65.2
Data 500, Upscale	94.05	80.77	88.51	66.81
Data 1000, Upscale	92.63	81.66	88.73	65.58

#### IV. Results and Discussions

본 장에서 COCO 데이터셋을 활용하여 제안한 기법의 실험 및 연구 결과를 나타낸다. 기존 사전학습 된 YOLOv5에 대하여 Super-Resolution 기법을 적용한 결과 다음 그림과 같이 객체 탐지 성능이 개선되었음을 확인하였다.



Fig. 1. Object detection results of YOLOv5 using the original COCO validation dataset (Left) and the proposed technique (Right)

그림은 COCO 데이터셋의 검증 데이터에 대하여 YOLOv5 모델의 객체 탐지 수행 결과를 일부 나타낸 것이다. 사람, 의자, 컵, 그릇을 포함한 38가지 객체가 검출되었다. 작은 객체에 대한 신뢰도가 대체로 낮으며, 테이블 위의 그릇을 피자라고 판단하는 것과 같은 오탐지를 확인할 수 있었다. 반면, 제안 기법을 적용한 YOLOv5 객체 탐지에서 식기 등과 같은 작은 물체에 탐지에 대해 이전보다 높은 수치를 보였다. 또한, 포크나 나이프 같은 작은 객체를 추가로 탐지하였으며 전과 같은 피자나 물병에 대한 오탐지를 확인할 수 없었다.

다음 그림 2는 COCO 데이터셋의 검증 데이터 중에서 YOLOv5의 오탐지와 제안 기법을 통해 이를 개선한 사례이다.



Fig. 2. Results showing the original image mistakenly identifying people as dogs (Left), and successfully corrected detections (Right) achieved by the proposed technique.

위 그림의 좌측은 제안 기법을 적용하기 전 객체 탐지 수행 결과이며, 테니스 공을 치는 사람에 대한 클래스를 개(dog)라고 잘못된 결과를 도출하였다. 반면, 제안 기법을 적용한 객체 탐지는 제대로 사람(person)이라고 판단하였다. 다음 표2는 연구 수행 결과이다.

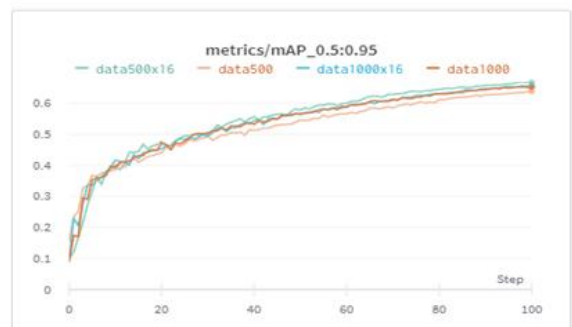
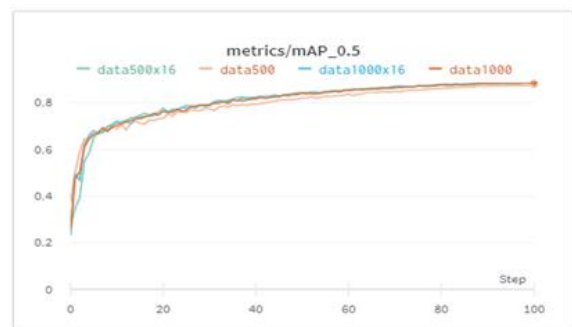
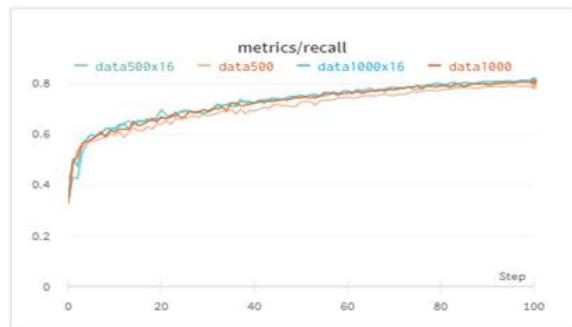
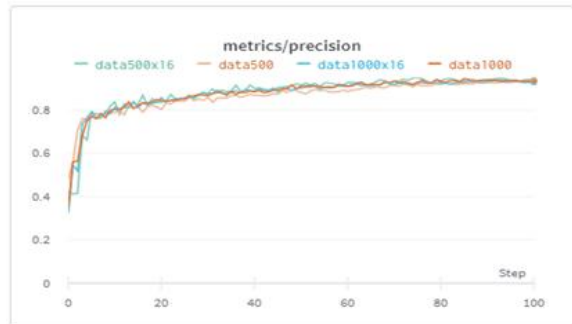


Fig. 3. Experiment result graph

데이터 증강 전 사람 클래스로 레이블링된 500개의 학습 데이터 모델의 mAP\_0.5와 mAP\_0.5:0.95는 각각 87.58, 63.79의 수치를 보였다. 제안 기법을 적용해 학습 데이터를 1,000개로 늘리고 테스트 및 검증 데이터에 대해 Super-Resolution 기법을 적용해 16배 확대시킨 모델은 mAP\_0.5와 mAP\_0.5:0.95에 대하여 각각 88.73과 65.68의 수치로 성능이 향상되었음을 확인하였다.

## V. Conclusions

딥러닝 기술의 발달로 여러 분야에 걸쳐 객체 탐지 기술이 활용되고 있다. 하지만 특정 상황에 대한 학습 데이터셋이 충분하지 않아 모델 성능이 떨어지거나, 풀링 및 업샘플링 과정에서 국소적인 정보가 소실되어 작은 개체의 탐지가 어렵다는 단점을 지목했다. 본 논문은 COCO 데이터셋에 대하여 데이터 증강기법을 통해 학습 데이터를 구성하여 이전에 비해 클래스가 부족한 가려진 객체 탐지 예측 정확도에 대한 모델 성능 향상을 보였고, Super-Resolution을 적용한 확대 기법을 제안하여 작은 객체 탐지에 관한 성능향상을 보여 유의미한 결과를 도출하였다. 이 같은 연구결과는 데이터셋 확보가 어려운 산업 및 안전 분야에서 위험을 조기 감지 하거나, 질병 진단과 예후 평가에 중요한 역할을 할 수 있을 것으로 기대된다.

## ACKNOWLEDGEMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 2022202090003C)

## REFERENCES

- [1] J.S. Lee et al, "Trends on Object Detection Techniques Based on Deep Learning," Electronics and Telecommunications Trends (ETRI), Vol. 33, No. 4, pp.23-23, 2018.
- [2] Kyoseong Hwang and Jungmok Ma, "Improvement of Object Detection Performance for Enemy Tanks Using the Combination of Data Augmentation Models," Korean Journal of Computational Design and Engineering, Vol. 27, No. 2, pp. 148-159. June 2022.
- [3] Ian J. Goodfellow et al, "Generative Adversarial Networks," In Proceedings of the Neural Information Processing Systems (NIPS), pp. 2672-2680, 2014.
- [4] Joseph Redmon et al, "You Only Look Once: Unified, Real-Time Object Detection," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.
- [5] Alec Radford, Luke Metz, Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," The International Conference on Learning Representations (ICLR), 2016.
- [6] Christian Ledig et al, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4681-4690, 2017.
- [7] Tsung-Yi Lin et al, "Microsoft COCO: Common Objects in Context," European Conference on Computer Vision (ECCV), pp. 740-755, 2014.