

데이터 증강 기반 Saliency U-Net을 이용한 저해상도 객체 검출 방법에 대한 연구

안명수*, 임혜연**, 강대성***

A Study on the Low Resolution Object Detection Method using Saliency U-Net and Data Augmentation

MingShou An*, Hye-Youn Lim**, and Dae-Seong Kang***

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.RS-2023-00247045)

요 약

저해상도 영상에서의 객체 검출은 영상의 품질이 낮아 산업적인 응용에서는 전통적인 객체 검출 알고리즘에만 의존할 시 잡음에 취약하여 성능 저하를 초래하는 경우가 많다. 이러한 문제를 해결하기 위하여 본 논문에서는 지식 증류 기반의 GAN(Generative Adversarial Networks) 알고리즘을 적용하여 고해상도 영상 데이터를 회전, 크기, 반전, 해상도 등의 변환을 거쳐 객체에 대한 특징 데이터를 증강 학습하여 데이터를 생성한다. 또한, 생성된 데이터에 대하여 Saliency 메커니즘 기반의 U-Net을 활용하여 특정 객체를 검출하고 추적하는 방법을 연구한다. 기존 검출 방법과의 비교 분석 실험을 통해 제안하는 방법이 저해상도 영상에서 더 높은 정확도와 강인성을 제공함을 검증하였다. 실험 접근 방식은 3개의 산업용 검출 분야 관련 공개 데이터 셋에서 성능 평가를 진행하였으며, 기존 방법에 비해 mAP가 5% 이상 개선되었음을 확인하였다.

Abstract

Object detection in low-resolution images is often vulnerable to noise and results in performance degradation when relying only on traditional object detection algorithms in industrial applications due to low image quality. In order to solve this problem in this paper, the knowledge distillation-based GAN (Generative Adversarial Networks) algorithm is applied to convert high-resolution image data such as rotation, size, inversion, resolution, etc. and generation. In addition, we study a method of detecting and tracking a specific object using U-Net based on the Saliency mechanism for the generated data. Through comparative analysis experiments with existing detection methods, it was verified that the proposed method provides higher accuracy and robustness in low-resolution images. As for the experimental approach, the performance was evaluated in three open data sets related to industrial detection, and it was confirmed that the mAP was improved by more than 5% compared to the existing method.

Keywords

knowledge distillation, GAN, U-Net, saliency mechanism, object detection

* 중국 서안공업대학교 컴퓨터공학과 강사
- ORCID: <https://orcid.org/0000-0002-1180-8916>
** 동아대학교 전자공학과 조교수
- ORCID: <https://orcid.org/0000-0002-5189-461X>
*** 동아대학교 전자공학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0003-0186-2430>

· Received: Apr. 18, 2023, Revised: Jul. 21, 2023, Accepted: Jul. 24, 2023
· Corresponding Author: Dae-Seong Kang
Dept. of Dong-A University, 37 NaKdong-Daero 550, beon-gil saha-gu, Busan, Korea
Tel.: +82-51-200-7710, Email: dskang@dau.ac.kr

I. 서론

본 논문에서는 기존의 객체 검출 방법들은 저해상도 영상에서 낮은 대비와 노이즈에 취약한 한계가 있다는 점을 보완하기 위해 희미한 특성 영상에서의 특징 검출에 강인한 딥러닝 방법을 제안하고자 한다. 저해상도 영상은 픽셀 수가 적거나 품질이 낮은 영상을 의미한다. 이러한 저해상도 영상은 디지털 확대, 데이터 압축, 노이즈, 낮은 대비 등의 원인[1]으로 인해 발생할 수 있으며, 저해상도 영상 처리는 컴퓨터 비전 분야에서 중요한 연구 주제로, 특히 희미한 영상에서의 특징 검출, 영상 향상, 영상 복원 등의 문제를 다룬다[2][3]. 딥러닝 기반 저해상도 영상 처리 방법은 다양한 네트워크 구조, 손실 함수, 최적화 기법을 사용하여 더 나은 결과를 달성할 수 있다.

현재 저해상도 영상 분야에서는 저해상도 영상의 품질을 향상 시켜 응용하는 연구가 가장 많이 진행되고 있다. 이 방면의 연구는 주로 3가지 방향으로 나뉜다. 첫째 초고해상도, 즉 저해상도 영상으로부터 변환을 통해 고해상도 영상을 획득할 수 있는 기술로, GAN, CycleGAN[4] 등 방법들이 크게 기여하고 있다. 둘째 영상보간, 즉 픽셀간의 빈 공간을 채워주는 방법으로 딥러닝 기반 영상 보간 알고리즘[5]이 있으며, 비교적 높은 성능을 보여주고 있다. 셋째 영상 복원 및 향상, 즉 저해상도 영상에서 노이즈, 왜곡, 밝기, 대비, 채도 등을 개선하여 원래의 고해상도 영상으로 복원하는 기술로, CNN[6], Autoencoder[7][8], U-Net[9][10] 등의 딥러닝 모델에 대한 많은 연구가 진행되고 있다.

본 연구에서는 희미한 저해상도 영상에서의 객체 검출 성능을 높이기 위해 기존 방법들의 한계를 극복하는 지식 증류와 GAN 구조[11]를 결합하여 고해상도의 데이터 셋을 기반으로 저해상도의 데이터 셋을 생성하는 데이터 증강 기술을 제안한다. 또한 저해상도 영상에서 특정 객체를 검출하기 위해 Saliency 메커니즘 기반의 U-Net[12]을 적용한 신경망 학습 모델[13]을 설계하고 테스트 데이터 셋에서의 효율성을 검증한다.

II. 관련 이론

2.1 데이터 증강

데이터 증강은 적은 양의 데이터를 바탕으로 딥러닝 알고리즘을 통해 데이터의 양을 늘려 검출, 분류, 추론 등 현실 문제를 해결하기 위한 학습 과정에서의 응용 분야의 데이터가 부족한 경우에 대비해서 효율적인 학습을 수행하여 더 강인한 딥러닝 모델을 구성하기 위해 사용되는 기법이다. 데이터 증강 모델에서 가장 많이 사용되고 있는 모델은 Ian Goodfellow가 제안한 GAN이다. GAN을 활용하면 수집 및 가공된 적은 양의 데이터 셋을 데이터 증강(Augmentation)을 통해 새로운 데이터 셋을 구축하여 학습에 활용할 수 있으며, 그림 1은 GAN의 구조와 동작을 나타낸 것이다.

2.2 지식 증류

지식 증류 기법은 이미 학습된 모델(교사네트워크, Teacher network)을 이용하여 학습이 부족한 모델(학생네트워크, Student network)을 학습시키는 방법이다. 이 방법은 학생모델에서의 확률 분포에 대한 예측 계산이 교사모델의 결과를 따라가므로 목표 성능까지 쉽게 도달 가능하다는 장점이 있다.

지식 증류 과정은 다음과 같다.

(1) 교사모델 학습: 큰 신경망을 학습시켜 높은 성능에 달성한다.

(2) 교사모델 소프트 라벨: 교사모델을 사용하여 학습 데이터 셋에 대한 예측을 수행하고, 출력 확률 분포를 부드럽게 만들어 관련 매개변수를 사용하여 소프트 라벨을 생성한다. 소프트 라벨은 교사모델의 예측 확률을 표현하므로, 원본 라벨보다 더 많은 정보를 포함한다.

(3) 학생모델 학습: 학생모델은 소프트 라벨을 목표로 하여 손실 함수를 최소화하도록 하는 과정을 수행하며, 교사모델의 성능을 그대로 복사하여 학습한다.

(4) 성능평가: 학습이 완료되면 학생모델의 성능을 평가하여 교사모델과 유사한 성능을 달성하면서도 더 적은 비용을 사용하게 된다.

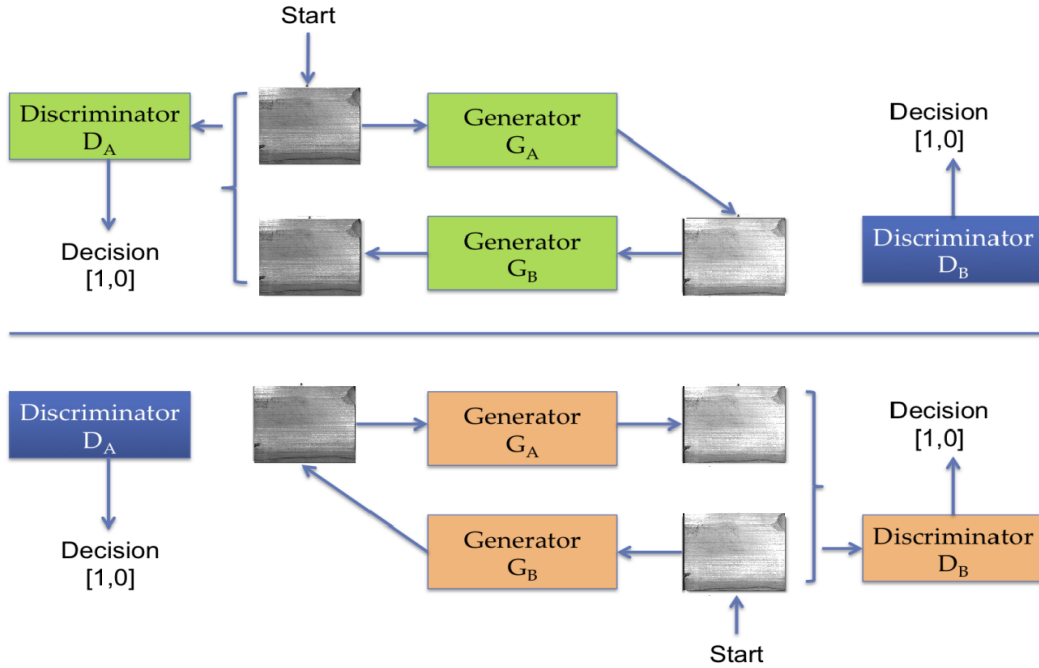


그림 1. GAN 신경망 Layer의 구조
Fig. 1. Layer structure of GAN

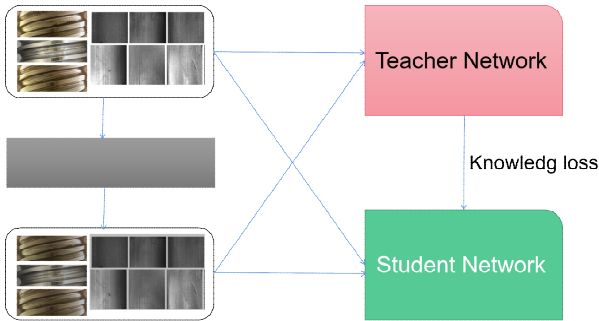


그림 2. 지식 증류 구조
Fig. 2. Structure of knowledge distillation

으로 인코딩 과정에서는 입력 영상의 특징을 추출할 수 있도록 채널 양을 확장하여 차원을 감소시키며, 디코딩 과정에서는 인코딩과정에서 획득한 저차원 정보만 이용하여 채널 양을 줄이고 차원을 확장하여 고차원의 영상으로 복원한다. 하지만 인코딩 과정에서 차원 감소를 거치면서 영상 내 객체에 대한 정확한 위치 정보가 손실되고, 디코딩 과정에서도 저차원의 정보만을 이용하기 때문에 손실 위치 정보를 회복하기는 쉽지 않다.

2.3 U-Net 기반 객체 검출

U-Net은 논문 'U-Net: Convolutional Networks for Biomedical Image Segmentation'에서 처음으로 제안되었으며, 기본 원리는 저차원 정보뿐만 아니라 고차원 정보까지 이용하여 영상의 특징을 추출함과 동시에 검출하고자 하는 객체의 정확한 위치 파악도 가능하게 하는 것이다. 이를 위해서는 인코딩 과정에서의 각 레이어에서 획득한 특징을 디코딩 과정의 각 레이어에 통합하는 방법을 기반으로 설계하며, 그림 3은 U-Net의 구조를 보여준다.

U-Net은 오토인코더(Autoencoder)와 같은 인코더-디코더(Encoder-decoder) 기반 모델에 속한다. 일반적

III. 제안하는 방법

3.1 지식 증류 기반 GAN 데이터 증강

기존의 지식 증류 기법은 전이학습에서 사전 학습 완료된 모델에 의존하여 재학습되기 때문에, 컴퓨팅 자원이나 환경이 제한되는 분야에서 적용하는데 발생하는 한계점을 극복하기 위하여 사전 학습 완료된 신경망 모델에서 다른 신경망 모델로 이전하여 학습을 수행하는 것이다. 지식 증류 학습방법은 유클리디안(Euclidean) 거리의 최소화를 위해 비용 함수(Cost function)를 설정하여 학습을 수행하게 된다.

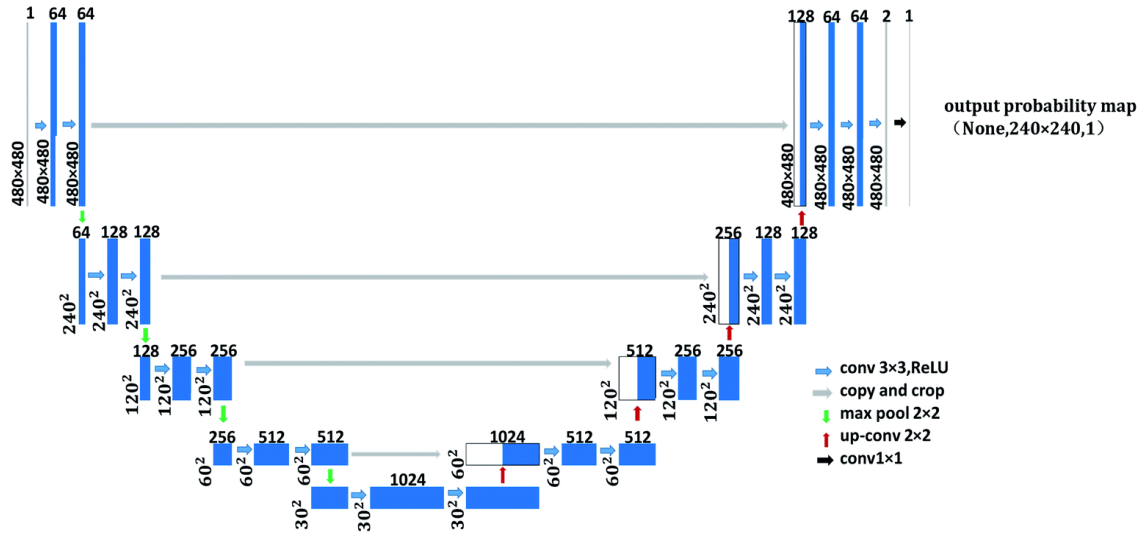


그림 3. U-Net 구조
Fig. 3. Structure of U-Net

이와 달리, 본 논문에서는 기존 방식과는 다른 복잡한 GAN 모델 내부의 지식을 표현하는 흐름정보를 추출하여 다른 대상 모델로 이전하여 학습을 수행하는 새로운 방식을 제안한다. 이는 생성자(Generator)를 학습 대상 네트워크로 설정하고, 상기 네트워크 모델에서 추출한 흐름정보들을 MLP(Multi-Layer Perceptron)의 판별자(Discriminator)가 사전 학습된 네트워크 모델을 통해 획득한 흐름정보들과 서로 비교하여 구분할 수 없을 때까지 적대적 학습과정을 거쳐 수행하는 방식을 적용한다. 학습과정은 다음과 같다. 교사모델에서 학생모델로 지식을 전달하는 지식 증류 과정을 수행하는데, 이때 학생 모델은 교사모델의 성능을 모방하면서도 더 적은 계산 비용을 요구하게 된다.

GAN의 생성자와 판별자 중 하나 또는 모두를 학습된 학생 모델로 설정하고, 설정된 학생모델은 고성능 교사모델로부터 지식을 증류했기 때문에 GAN 학습의 초기 성능이 높아지게 된다. 이러한 접근 방식은 이미지 생성, 이미지 변환, 데이터 생성 등 다양한 응용으로 활용할 수 있다. 그림 4는 지식증류 기반의 GAN 구조를 보여주고 있다.

3.2 SU-Net 학습 모델

본 논문에서 제안하는 Saliency U-Net은 U-Net 구조를 기반으로 시각적 주목성(Saliency) 또는 관심영역(Region of interest)의 예측을 위한 신경망이다. Saliency는 영상에서 인지적으로 중요한 부분을 감지하는 작업으로, 영상 처리 및 컴퓨터 비전 분야에서 다양한 응용이 가능하다. U-Net은 합성곱 신경망(CNN)의 변형으로, 영상 분할(Image segmentation) 작업에 특화된 구조를 가지고 있어, 단순히 윤곽 검출뿐 아니라, 픽셀단위로 객체에 대한 정보를 검출 및 분석 가능하다.

U-Net은 인코더-디코더 구조를 가지며, 인코더에서 영상의 고수준 특징을 추출하고, 디코더에서 고해상도 분할 맵(Segmentation map)을 생성한다. 이 과정에서 U-Net은 스킵 연결(Skip connections)을 사용하여 인코더와 디코더 사이의 정보를 공유함으로써, 세부 정보를 보존하고 효과적인 분할을 수행한다.

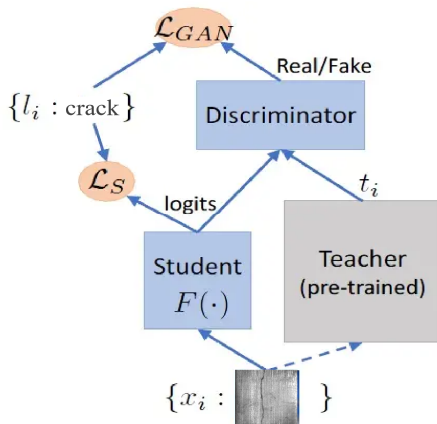


그림 4. 지식 증류 기반 GAN 구조
Fig. 4. Structure of GAN based on knowledge distillation

Saliency U-Net은 기본 U-Net 구조를 활용하여 시각적 주목성 맵(Saliency map)을 생성하는 모델이다. 이 모델은 영상을 입력으로 받아, 각 픽셀이 얼마나 주목할만한지에 대한 확률 분포를 출력한다. 일반적으로 마지막 레이어에서 시그모이드(Sigmoid) 활성화 함수를 사용하여 각 픽셀의 주목성 값이 0과 1 사이의 범위를 가지도록 한다. 이렇게 생성된 saliency 맵은 영상 내 주요 객체나 특징을 감지하는데 사용될 수 있다. 그림 5는 saliency U-Net 모델의 구조를 보여준다.

IV. 실험 결과

4.1 지식 증류 기반 GAN 모델 분석 실험

본 논문에서는 데이터 확보를 위해 소량의 학습 데이터를 인위적으로 변화시켜 새로운 학습 데이터를 대량 확보 가능한 방법론을 적용하여 저해상도 영상 데이터 셋을 구성한다. 저해상도로의 변환은 단순한 다운샘플링 방법으로도 획득 가능하지만, 이렇게 생성된 데이터는 변환과정에서 영상이 갖고 있던 많은 정보를 손실하게 된다.

본 논문에서는 지식 증류(Knowledge distillation) 기반 GAN을 사용한 데이터 증강 기법을 적용하여 고해상도 영상 데이터셋을 기반으로 저해상도 영상 데이터셋을 구성하는 방법을 사용한다. 제안하는 방

법은 학습 수렴속도 및 학습에 소요되는 시간을 단축시킬 수 있으며 또한 영상 정보 손실을 최대한 줄일 수 있는 것이 장점이다. 실험에서는 먼저 GAN 모델을 적용하여 Generator는 저해상도 영상을 입력하여 고해상도 영상을 생성하며, Discriminator 과정에서는 생성된 고해상도 영상을 실제 고해상도 영상과 구별하여, 학습된 고해상도 모델의 정보를 저해상도 모델로 전달한다.

하지만 지식 증류를 수행할 때 사전 학습 네트워크를 사용하기 때문에 구조 측면에서 뉴런 크기 등 일치 여부를 확인하고, 학습률 조절을 통해 학습 곡선의 왜곡을 감소시킬 수 있다. 실험에서 사용된 데이터 셋은 산업현장용 영상 데이터 셋을 이용하여 저해상도 영상 데이터 셋으로 생성하여 최종 Saliency U-Net을 기반으로 객체 검출을 수행한다. 표 1은 실험에서 사용된 데이터 셋들이다.

표 1. 데이터 셋 리스트
Table 1. List of datasets

| Dataset list | |
|--------------|---|
| 1 | Industrial wear forecasting dataset 2021 https://github.com/20obe/BSData |
| 2 | Magnetic tile dataset 2018 https://github.com/abin24/Magnetic-tile-defect-datasets . |
| 3 | Textures classification dataset 2019 https://github.com/abin24/Textures-Dataset |

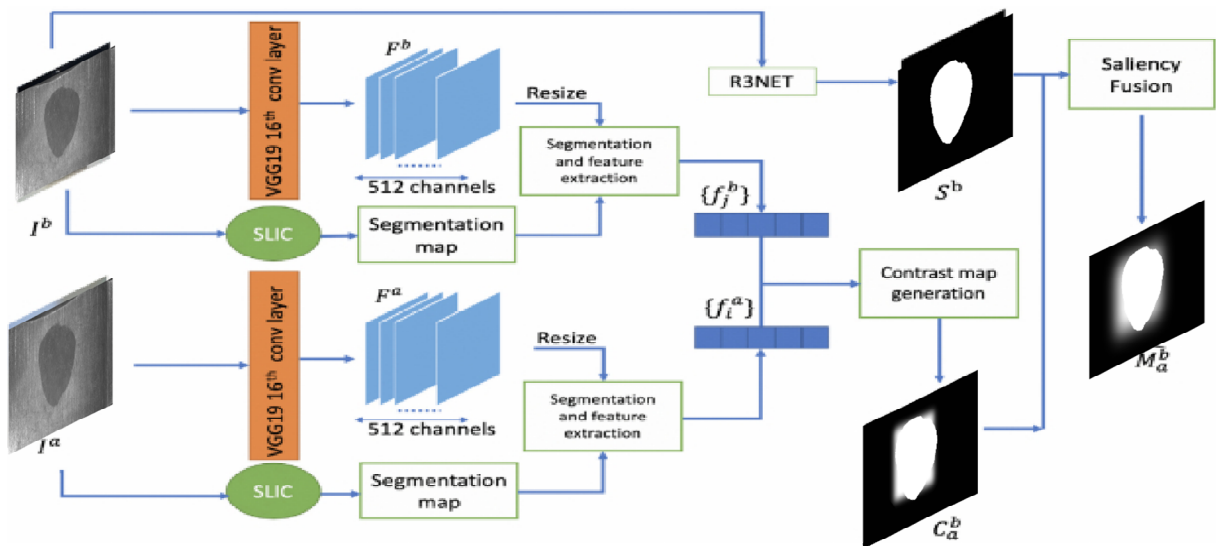


그림 5. Saliency U-Net 구조
Fig. 5. Structure of saliency U-Net

제안 방법의 데이터 생성 검증을 위해 사용된 사전 학습 모델로는 PG-GAN(Generator)과 DC-GAN(Discriminator)을 사용한다. GAN 기반 데이터 증강 신경망 모델의 학습 수렴은 Generator와 Discriminator Loss의 변화를 통해 평가할 수 있다. 표 2는 지식 증류를 적용하기 전과 후의 GAN 모듈을 100회와 1,000회 Epoch에서 학습한 객체 영역의 정확도를 나타낸 결과이다.

그림 6은 지식 증류의 적용 전과 후의 GAN 모듈에서 mAP 0.5IoU에 수렴하기까지의 학습 곡선을 나타낸다.

표 2. Epoch=100, 1,000으로 학습했을 때의 mAP 결과
Table 2. Results of learning with Epoch=100 and 1,000

| Methods | Epoch | mAP(%) |
|-----------------|-------|--------|
| GAN | 100 | 32.32% |
| | 1,000 | 34.33% |
| CycleGAN | 100 | 31.87% |
| | 1,000 | 38.32% |
| proposal method | 100 | 41.22% |
| | 1,000 | 42.12% |

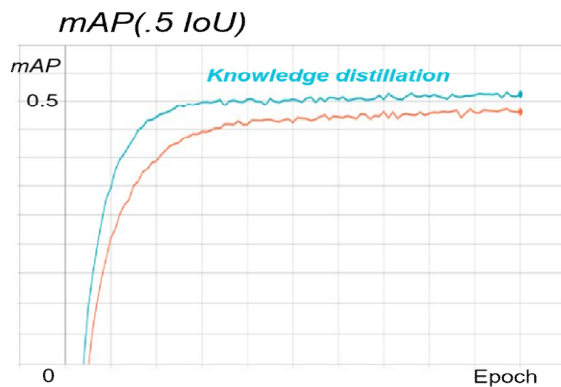


그림 6. IoU 0.5에 수렴하기까지의 학습 곡선
Fig. 6. Learning convergence curves at IoU 0.5

4.2 Saliency U-Net 기반 객체 검출 결과

실험에서는 표 1에 명시된 데이터셋에서 증류 기반 GAN 모델을 통해 생성된 저해상도 영상에 대해 Saliency U-Net을 이용하여 객체 영역을 학습시켰다. 테스트 영상으로는 각 데이터셋의 일부를 추출하여 사용하였으며, 기존 대표적인 모델들과 정확도 비교 실험을 진행하였다.

표 3은 기존 모델과 제안 모델간의 정확도 비교 결과를 나타낸 것이다. 데이터 셋 1에서의 정확도가 5% 이상으로 가장 많이 향상된 것을 확인할 수 있었다. 그림 7은 데이터 셋 1에서의 크랙, 블로 홀, 브레이크에 대하여 Faster R-CNN, VGG-19, YOLO v5, Saliency U-Net을 활용하였을 때의 검출 결과를 보여준다.

표 3. 모델간 정확도 비교 결과
Table 3. Accuracy comparison results of each model

| Models | ACC(%) | | |
|----------------|----------|----------|----------|
| | dataset1 | dataset2 | dataset3 |
| Faster R-CNN | 79 | 82 | 84 |
| VGG-19 | 83 | 84 | 82 |
| YOLO v5 | 86 | 91 | 94 |
| Saliency U-Net | 91 | 93 | 93 |

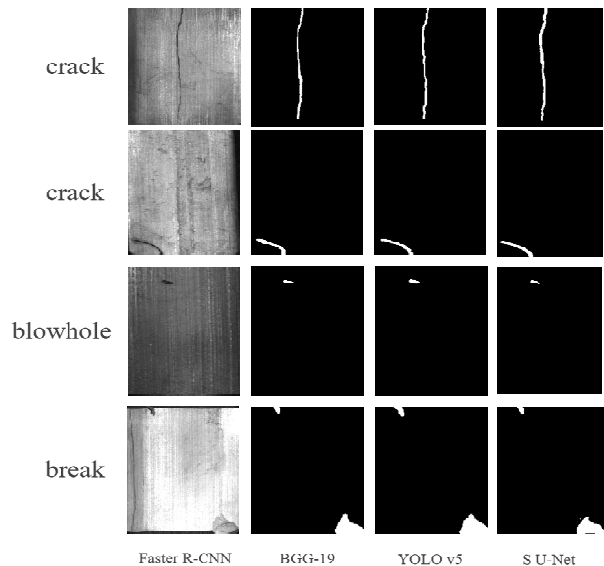


그림 7. 데이터 셋 1에서의 검출 결과
Fig. 7. Detection results in dataset 1

V. 결 론

낮은 해상도의 영상에서 객체를 검출하고 분석하기에는 기존 방법들이 취약한 경우가 많다. 이에 본 논문에서는 먼저 지식 증류 기반의 GAN 알고리즘을 적용하여 고해상도의 영상으로부터 저해상도의 영상을 생성하는 학습 데이터 확보 방법을 제안한다. 다음 생성된 데이터에 대해 Saliency U-Net 알고리즘을 활용한 객체 검출 알고리즘을 제안하였다.

실험에서는 지식 증류 기반 GAN알고리즘과 GAN, CycleGAN 알고리즘을 각각 Epoch가 100일때와 1,000일 때의 mAP 값을 비교한 결과 제안하는 방법이 기존 방법에 비해 최소 4%, 최대는 10% 가까이 개선 된 것을 확인할 수 있었다. 또한 정확한 객체 검출 결과를 분석하기 위해 Saliency U-Net 방법과 기존 대표적인 알고리즘을 통한 검출 결과를 비교하였으며, 제안하는 방법이 기존 방법에 비해 정확도가 5% 향상되었음을 확인할 수 있었다.

추후 연구에서는 학습 모델의 연산량을 줄이기 위한 손실 함수의 변환 및 조합을 통해 계산량이 적은 모델로 대체할 수 있는 함수로 수정하여 연산량을 줄이는 방법을 모색하고자 한다.

References

- [1] H. Chen, Y. Pei, H. Zhao, and Y. Huang, "Super-resolution guided knowledge distillation for low-resolution image classification", *Pattern Recognition Letters*, Vol. 155, pp. 62-68, Mar. 2022. <https://doi.org/10.1016/j.patrec.2022.02.006>.
- [2] W. Liu, S. Liao, W. Ren, W. Hu, and Y. Yu, "High-level semantic feature detection: A new perspective for pedestrian detection", *Proc. of the IEEE/CVF conference on computer vision and pattern recognition(CVPR)*, pp. 5187-5196, Jun. 2019.
- [3] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration", *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, pp. 5728-5729, Jun. 2022.
- [4] Y. Li, H. Wang, and X. Dong, "The denoising of desert seismic data based on cycle-GAN with unpaired data training", *IEEE Geoscience and Remote Sensing Letters*, Vol. 18, No. 11, pp. 2016-2020, Nov. 2021. <https://doi.org/10.1109/LGRS.2020.3011130>.
- [5] C. H. Neetha, C. J. Moses, and D. Selvathi, "Image Interpolation Using Non-adaptive Scaling Algorithms for Multimedia Applications—A Survey", *Advances in Automation, Signal Processing, Instrumentation and Control: Select Proceedings of i-CASIC*, Springer Singapore, Vol. 700, pp. 1509-1516, Mar. 2021. https://doi.org/10.1007/978-981-15-8221-9_141.
- [6] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 173, pp. 24-49, Mar. 2021. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>.
- [7] J. E. Jeong, H. J. Kim, and J. H. Chun, "Automatic Augmentation Technique of an Autoencoder-based Numerical Training Data", *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, Vol. 22, No. 5, pp. 75-86, Oct. 2022.
- [8] G. Zhang, Y. Liu, and X. Jin, "A survey of autoencoder-based recommender systems", *Frontiers of Computer Science*, Vol. 14, pp. 430-450, Aug. 2019. <https://doi.org/10.1007/s11704-018-8052-6>.
- [9] T. H. Kim and J. J. Park, "Image Segmentation for Fire Prediction using Deep Learning", *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, Vol. 23, No. 1, pp. 65-70, Feb. 2023.
- [10] K. S. Lee, W. M. Gal, and M. J. Lim, "Sampling-based Super Resolution U-net for Pattern Expression of Local Areas", *The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)*, Vol. 22, No. 5, pp. 185-191, Oct. 2022.
- [11] H.-Y. Lim, J.-M. Lee, and D.-S. Kang, "A Method for Improving Learning Convergence Curve and Learning Time of DA-FSL Model using Knowledge Distillation", *Journal of KIIT*, Vol. 18, No. 10, pp. 25-32, Oct. 2020. <https://doi.org/10.14801/jkiit.2020.18.10.25>.

- [12] X. Gong, L. Qingge, Q. Liu, and P. Yang, "Improved U-Net-Like Network for Visual Saliency Detection Based on Pyramid Feature Attention", *Wireless Communications and Mobile Computing*, Vol. 2022, Aug. 2022. <https://doi.org/10.1155/2022/1108462>.
- [13] Z. Ning, S. Zhong, Q. Feng, W. Chen, and Y. Zhang, "SMU-net: saliency-guided morphology-aware U-net for breast lesion segmentation in ultrasound image", *IEEE transactions on medical imaging*, Vol. 41, No. 2, pp. 476-490, Feb. 2022. <https://doi.org/10.1109/TMI.2021.3116087>.

저자소개

안 명 수 (MingShou An)



2016년: 동아대학교
전자공학과(공학박사)
2017년 ~ 현재 : 중국 서안공업
대학교 컴퓨터공학과 강사
관심분야 : 영상처리, 패턴인식,
인공지능

임 혜 연 (Hye-Youn Lim)



2013년 : 동아대학교
전자공학과(공학박사)
2013년 ~ 현재 : 동아대학교
전자공학과 조교수
관심분야 : 영상처리, 패턴인식,
인공지능

강 대 성 (Dae-Seong Kang)



1994년 : Texas A&M 대학교
전자공학과(공학박사)
1995년 ~ 현재 : 동아대학교
전자공학과 교수
관심분야 : 영상처리, 패턴인식,
인공지능