논문 2023-1-9 http://dx.doi.org/10.29056/jsav.2023.3.09

고해상도 드론 데이터에 이미지 슬라이싱과 Transformer 기법을 적용한 불법작물 객체 탐지 시스템 설계 및 구현

김현수*, 이예슬**, 신동명**, 이찬재***, 김명호*†

Design and Implementation of an Illegal Crop Object Detection System using Image Slicing and Transformer Techniques to High-Resolution Drone Data

Hyun-Soo Kim*, Ye-Seul Lee**, Dong-Myung Shin**, Chan-jae Lee***, Myung-Ho Kim*†

요 약

본 논문에서는 고해상도 드론 데이터를 활용하여, 4K, 8K의 이미지에 슬라이싱 기법을 적용하는 방안에 대해 연구하고, 그를 활용하여 개인정보 객체(차, 사람)를 탐지하고 비식별화하여, 비식별화된 이미지에서 불법작물 객체를 탐지하고 시각화하는 시스템에 대해 설계하고 구현하였다. 개인정보 객체탐지 및 비식별화를 위해, Single-Stage 기법인 Yolov5와 Gaussian Blurring 기법을 적용하고, 불법작물 객체탐지를 위해 Swin Transformer와 Soft-Teacher, Faster-RCNN 기법을 적용하였으며, 이미지 슬라이싱 기법으로는 SAHI 오픈소스 프레임워크를 사용하였다. 불법작물 객체탐지 모델은 Backbone 네트워크로 Swin Transformer, 검출기로 Faster-RCNN을 사용하는 앙상블된 Soft-Teacher 모델을 사용하였다. 이 모델에 이미지 슬라이싱 기법을 적용하여 실험한 결과 mAP가 0.663으로 기존의 이미지 슬라이싱 기법을 적용하지 않은 모델의 mAP인 0.456보다 향상된 것을 확인할 수 있었다.

Abstract

In this paper, we studied how to apply slicing techniques to 4K and 8K images using high-resolution drone data, and used them to detect and de-identify personal information objects (car, people), and designed and implemented a system that detects and visualizes illegal crop objects in de-identified images. For privacy object detection and de-identification, single-stage techniques such as Yolov5 and Gaussian Blurring were applied, SwinTransformer, Soft-Teacher, and Fast-RCNN techniques were applied to detect illegal crop objects, and SAHI open source framework was used as image slicing techniques. The illegal crop object detection model used an ensemble Soft-Teacher model using SwinTransformer as a Backbone network and Fast-RCNN as a detector. Experiments by applying image slicing techniques to this model showed that the mAP was 0.663, which is improved from 0.456, which is the mAP of the model without applying the conventional image slicing techniques.

한글키워드 : 객체 탐지, 이미지 슬라이싱, 인공지능, 딥러닝, 드론

keywords : object detection, image slicing, artificial intelligence, deep-learning, drone

* 숭실대학교 AI,SW 융합학과
** 엘에스웨어㈜
*** 오케스트로㈜
† 교신저자: 김명호(email: kmh@ssu.ac.kr)
접수일자: 2023.03.10. 심사완료: 2023.03.17.
게재확정: 2023.03.20.

1. 서 론

최근 드론은 자율주행, 인공지능, 사물인터넷 등 미래 기술과 융합돼 다양한 산업에 응용되며, 드론이 생성하는 대용량 데이터를 빠르고 정확하게 처리하여 인공지능 학습 및 분석에 활용되고 있다. 최근 인공지능 학습 및 추론 데이터는 AIHub(https://aihub.or.kr)와 같은 공공포털을 통해 공개되는데, 드론으로 촬영한 데이터의 경우 사람 얼굴 또는 자동차 번호판과 같은 개인정보를 포함할 경우 개인정보보호법에 의해 문제가 될 수 있다. 이에 미국, 영국 등 선도국에서는 개인정보를 보호하면서 빅데이터 활용을 위해 개인정보 비식별화 활용을 권고하고 있다. 우리나라는 빅데이터 활용 노하우가 부족한 상황으로 비식별화를 기반으로 한 빅데이터 활용과 도입 확산이 필요한 시점이다.

또한, 드론이나 고화질 카메라로 촬영한 고해상도 데이터를 기반으로 한 객체탐지 알고리즘도 꾸준히 연구되고 있다. 하지만 기존의 객체탐지 알고리즘은 640x360, 640x480과 같은 저해상도 이미지에서 객체를 추출하며, 검출되는 객체는 이미지의 약 60% 영역을 차지하고 있다. 하지만 고해상도 이미지에서 검출할 객체는 적은 영역에 표현되기 때문에 기존의 알고리즘으로는 정확하게 검출할 수 없어, 지속적인 연구가 필요하다.

이에 본 연구에서는 드론으로 촬영한 고해상도 데이터에서 인공지능 기법을 활용하여 고해상도 이미지를 효율적으로 처리하기 위해 이미지 슬라이싱 기법을 활용하고 사람, 차 객체를 탐지하고 자동으로 블러링하여, 데이터 공개 시에도 개인정보보호법에 위반되지 않는 학습데이터를 생성한다. 개인정보가 비식별화된 학습데이터 확장 및 추론 전 과정에 이미지 슬라이싱 기법을 적용하여 고해상도 데이터에서 불법작물 객체를 탐지하는 시스템을 제안한다.

2. 관련 연구

2.1 객체탐지 알고리즘

2.1.1. YOLO

YOLO는 이미지 픽셀, bounding box, coordinate, class probability 모두 single regression probelm으로 재구성해 이미지를 한 번 보는 것으로 그 안에 어떤 물체들이 있는지 어디에 있는지 파악할 수 있는 알고리즘이다[1]. single convolutional network가 이미지 전체에서 multiple bounding box에 대한 class probabiltity를 계산하는 통합된 모델을 사용한다.

2.1.2. Mask-RCNN

Mask-RCNN은 Faster-RCNN을 개선한 알고리즘으로 기존 Faster-RCNN의 Mask branch 추가, FPN(Feature Pyramid Network) 추가, Rolpooling을 RoI Align으로 대체와 같은 세 가지 문제점을 개선하였다[2]. 그 결과 다양한 객체 검출 시, 중간 정보를 많이 유실하는 Faster-RCNN보다 효율적으로 객체를 탐지한다.

2.1.3. TPH-YOLOv5

TPH-YOLOv5는 기존의 YOLOv5 모델의 예측 헤드(TPH)를 Transformer 방식으로 교체한 객체탐지 모델이다[3]. 또한 밀도가 높은 이미지에서 객체 탐지율을 높이기 위해 CBAM(Convolution Block Attention Module)을 사용한다. 이 결과 VisDrone2021 데이터셋을 기준으로 YOLOv5 모델보다 성능이 향상되었다.

2.1.4. RRNet

RRNet은 드론으로 촬영한 이미지에서 작은 개체를 탐지하는데 효율적인 알고리즘이다[4]. 이 알고리즘은 Anchor-Free 기반 검출기를 재회귀 모듈과 혼합하여 검출기를 구성하고 Resampling

기법을 사용하여 Multi-Scale 객체 탐지 분야에서 성능을 향상시켰다.

2.2 학습 데이터 생성 알고리즘

2.2.1. CycleGAN

CycleGAN은 2개의 생성자와 2개의 검출기의 학습을 통한 이미지를 타겟 이미지로 변환시켜주는 모델로 Image Translation의 대표적인 기술이다[5]. 이미지 생성자로써 Encoder-Decoder 구조인 U-Net과 ResNet을 사용하였다. U-Net 알고리즘은 극단적인 병목구조 모델로 학습이 불안정하고, ResNet은 적은 파라미터를 갖고 있어 학습은 안정적이라 결과 이미지의 모양 변화가 없다.

2.2.2. Soft-Teacher + SWIN-L

Soft-Teacher + SWIN-L은 적은 라벨링 데이터로 학습 성능을 향상시키는 연구로, Teacher 모델과 Student 모델을 사용하며 Backbone 네트워크로 Transformer 계열의 Swin-L을 사용한다[6,7]. 이 알고리즘은 Pseudo Label 기반 준지도 학습 모델로 기존 Transformer 계열 모델의 Intractable 문제 해결을 위해 계층 Feature Map과 Shifted Window Block을 적용하였다.

2.3 비식별화 알고리즘

비식별화 기술은 정보의 일부 또는 전부를 삭제하거나 대체하여 다른 정보와 결합하지 못하게 하는 기술을 의미하며, 이를 통해 특정한 정보를 알아볼 수 없게할 수 있다. 이미지 처리나 컴퓨터 비전에 사용되는 기본적인 변형 방법으로 이미지 필터의 합성(2D convolution)을 통해 이뤄진다. 즉, 원본 이미지보다 크기가 작은 필터를 이미지 전체에 걸쳐 이동하여 원본 이미지의 픽셀 값을 바꾼다. 블러링은 노이즈를 제거하는데 유용한 방법이며, 이미지를 좀 더 매끈하게 보이도록 만드는 효과를 내지만, 반대로 이미지의 특정 부분에 노이즈를 만들어 비식별화하는 데에 효과적이다.

표 1. 이미지 블러링 종류[8]
Table 1. Image Blurring Types

블러링 종류	특징
Averaging	필터 박스 내의 평균을 이용
Gaussain Blurring	2차원 가우시안 분포를 이용 박스 내dml 픽셀값을 가우시안 분포의 가중치로 평균을 내어 현재 픽셀값을 업데이트 가중치는 중심 픽셀에서 멀어질수록 급격하게 감소 중심 픽셀의 값은 주변의 픽셀들에 가장 크게 영향
Median Blurring	중간값을 이용한 블러링 현재 픽셀값을 박스 내의 픽셀값들의 중간값으로 대체함 박스 내 픽셀값들을 크기순으로 나열하여 중간값을 고르고, 이것을 현재 픽셀값 대체
Bilateral Riltering	노이즈를 제거하지만 이미지 내의 edge를 유지 이미지나 필터의 사이즈가 커지면 처리가 다소 느려지는 단점

2.4 이미지 슬라이싱 알고리즘

최근 컴퓨터 비전 분야에서는 객체 탐지 모델과 관련된 연구가 많이 진행되고 있다. 하지만 이에 사용되는 데이터셋은 Coco-Dataset과 같이 640x360의 저해상도 데이터를 사용하고 있다. 이에 드론으로 촬영되는 4K, 8K의 고해상도 데이터를 지금까지 연구된 객체탐지 모델에 적용할 경우, 이미지를 다운샘플링하여 활용해야하기 때문에 고해상도 이미지의 손실이 생긴다. 최근에는 작은 객체 탐지 알고리즘으로 MD-FPN, MD-IPN과 같은 새로운 Backbone 네트워크가 제안되고 있으며 이미지 슬라이싱 기법을 연구하

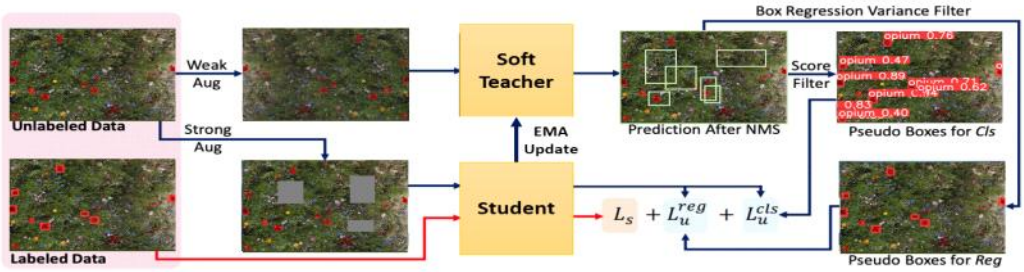

그림 1. 불법작물 객체탐지 모델
Fig 1. Illegal crop object detection model

고 제공하는 오픈소스 프레임워크도 이뤄지고 있다[9].

3. 시스템 설계 및 구현

3.1. 개인정보 객체탐지 모델

드론 촬영 데이터 내의 개인정보 객체 탐지를 위한 시스템의 시나리오 및 구성도는 그림 2와 같다.

본 시스템은 드론으로 촬영한 고해상도 이미지 데이터와 라벨링 데이터를 사용하여 사람 얼굴 및 자동차 번호판과 같은 개인정보 객체를 탐지하는 인공지능 모델을 학습하고, 학습된 모델은 고해상도 데이터 내의 개인정보 객체를 탐지하여 비식별화 한다[10,11].

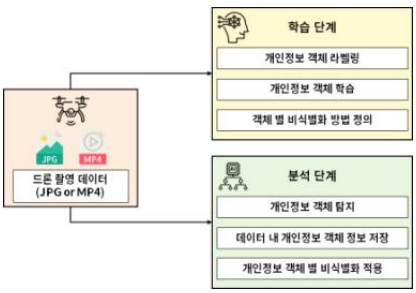

그림 2. 이미지 비식별화 프로세스
Fig 2. Image De-identification Process

개인정보 객체탐지에는 Yolov5 모델을 사용하여 개인정보 객체를 탐지하고, 탐지된 객체에 대하여 Gaussian Blurring을 적용하여 비식별화를 진행한다. 비식별화된 이미지는 원본과 동일한 해상도를 유지하며 Single-Stage 객체탐지 기법으로 데이터 전처리의 속도를 최소화하였다.

3.2. 불법작물 객체탐지 모델

고해상도 데이터에서 불법작물 객체를 탐지하기 위해 Backbone Network로 Swin Transformer를 사용하고, 라벨링된 데이터의 불균형을 보완하기 위해 Soft-Teacher 기법을 적용하였으며 구성은 그림 1과 같다.

기존 Soft-Teacher 모델은 Backbone 네트워크로 ResNet 네트워크 또는 Swin-L 네트워크를 사용하고 검출기로 Faster-RCNN 또는 HTC++를 사용였지만, 본 연구의 학습 모델은 Backbone 네트워크에 Transformer 계열인 Swin-T(Swin Transformer) 네트워크를 사용하고 검출기로 Faster-RCNN을 사용하여 불법작물 객체를 탐지한다. Swin-T 네트워크는 최근 객체탐지 분야에 적용할 경우 높은 성능향상을 보여 많이 적용되고 있는 방법으로, Shifted-Window 구조와 Feature Pyramid Network와 같은 계층 구조를 특징으로 한다.

3.3. 이미지 슬라이싱 기법

오픈 프레임워크인 SAHI(Slicing Aided Hyper Inference)는 이미지 내의 소규모 객체 탐지 성능 향상을 위한 연구로, 객체 탐지 AI 모델 학습 및 분석의 전 과정에서 이미지 슬라이싱 기법을 적용하는 오픈소스 프레임워크이다. AI 모델 학습 시, SF(Slicing aided Fine-tuning)를 적용하여 이미지를 슬라이싱하여 라벨링된 부분을 리사이즈한 뒤, 하나의 새로운 학습데이터로 사용한다.

AI 모델 추론 시, SAHI(Slicing Aided Hyper Inference)를 적용하여 하나의 이미지 학습 시, Input으로 슬라이싱 된 이미지들을 사용하고, Output으로 슬라이싱된 이미지를 하나의 이미지로 조합한다. 조합된 이미지 내의 검출된 객체의 중복을 제거하기 위해 NMS(Non-Maximum Suppression) 기법을 적용하여 중복된 부분의 Bounding-Box중 IoU가 가장 높은 것을 선택한다. 또한 슬라이싱된 이미지의 엣지 부분 손실을 최소화하기 위해, 엣지 부분은 중복하여 슬라이싱에 반영한다. 이에 본 연구에서는 Input 1으로 8K 이미지를 사용하고 SAHI는 이 Input 1을 사용하여 N개의 이미지로 슬라이싱하고 객체탐지 인공지능 모델의 Input 2로 전달한다. 인공지능 모델은 N개의 Input 2에 대해 추론하여 객체를 탐지하고 N개의 Output 1을 생성한다. SAHI는 N개의 Output 1을 다시 하나의 Output 2로 취합하여 원본과 동일한 해상도의 객체가 탐지된 이미지를 생성한다.

3.4. 웹 기반 통합 시스템

데이터 업로드, 개인정보 객체 탐지 및 비식별화, 불법작물 객체 탐지 결과 시각화를 위해 웹 기반 통합 시스템을 개발하였다. 본 시스템은 REST API를 사용하여 개인정보 객체 탐지부터 불법작물 객체 탐지까지 가능하며, 분석 결과를 그림 3과 같이 웹으로 확인할 수 있다. 또한 현재까지 수집된 데이터는 누구나 활용 가능하도록 공개데이터로 제공하고 있다.

그림 3. 웹 기반 통합 시스템 사용 불법작물 객체탐지 결과
Fig 3. Illegal crop object detection result using web based integrated system

4. 실험 및 평가

4.1. 데이터셋

드론으로 직접 수집한 8K 데이터셋 총 11,336장을 데이터셋으로 사용하였으며, 이 중 라벨링 된 데이터는 4,389장이다. 라벨링 된 이미지는 Train : Validation : Test = 4 : 1 : 2로 분류하여 학습 및 테스트에 사용하였으며, 비라벨링 데이터 6,947장은 Soft-Teacher 모델을 사용하여 준지도학습을 진행하였다.

4.2. 모델 성능 평가

모델 성능은 총 n가지 방법으로 실험을 진행하였다. 첫 번째 방법은 이미지 슬라이싱 기법을 학습 및 추론에 적용하지 않는 방법이다. 두 번째 방법은 추론 과정에만 이미지 슬라이싱 기법을 적용하는 방법이다. 세 번째 방법은 학습 및 추론 과정에 이미지 슬라이싱 기법을 적용하는 방법이다. 또한 불법작물 객체 탐지에 최적화된 모델을 만들기 위해 모델의 ROI head, anchor

sclae과 ratio, 이미지 슬라이싱 크기를 변경하였다. 또한 Fine-Tuning 과정에서 Learning late를 0.0001로 변경하였는데, 이를 변경하지 않는 경우 발산한다.

이미지 슬라이싱 기법을 객체탐지 모델 학습 및 추론 전 과정에 적용할 경우, mAP는 0.663으로 적용하지 않을 경우의 0.456보다 높은 수치로 성능이 향상되었다. 또한 불법작물 탐지 모델의 불법작물 탐지 정확도는 약 92%의 높은 정확도를 보였다.

표 2. 이미지 슬라이싱 기법 적용에 따른 모델 성능 측정
Table 2. Measurement of model performance according to application of image slicing technique

	이미지 슬라이싱 적용 여부		mAP@IoU=50
	학습	추론	
실험 1	X	X	0.456
실험 2	X	O	0.587
실험 3	O	O	0.663

5. 결 론

본 연구에서는 고해상도 이미지에서 개인정보 객체를 탐지 및 비식별화하고, 비식별화된 고해상도 이미지를슬라이싱하여, 슬라이싱 된 이미지에서 인공지능 모델이 객체를 탐지하고 결과를 시각화하는 시스템을 제시한다.

이 과정에서 8K 고해상도 이미지를 원본 그대로 사용하는 것보다 슬라이싱하여 사용하는 것이 모델의 객체탐지 정확도 향상에 도움이 되고, Backbone 네트워크로 SWIN Transformer를 사용하고 검출기로 Faster-RCNN을 사용하는 앙상블된 Soft-Teacher 모델이 성능이 좋다는 것을 보여주었다. 또한 개인정보 객체를 탐지하여 Gaussian Blurring 기법을 적용하면 개인정보보호법에 위반되지 않는 공개 가능한 학습데이터를 생성할 수 있음을 보여주었다.

본 연구에서 Backbone 네트워크로 Swin-T를 적용할 경우 ResNet을 사용할 경우보다 학습시간이 증가하는 것을 확인할 수 있었다. 향후 학습시간이 증가하는 부분을 개선하고 고해상도 이미지를 처리하는 경량화 모델에 대한 연구를 더 진행할 예정이다.

참 고 문 헌

[1] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." arXiv preprint arXiv:2207.02696 (2022).

[2] Ahmed, Belal, T. Aaron Gulliver, and Saif alZahir. "Image splicing detection using mask-RCNN." Signal, Image and Video Processing 14.5 (2020): 1035-1042.

[3] Zhu, Xingkui, et al. "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[4] Chen, Changrui, et al. "Rrnet: A hybrid detector for object detection in drone-captured images." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.

[5] Almahairi, Amjad, et al. "Augmented cyclegan: Learning many-to-many mappings from unpaired data." International Conference on Machine Learning. PMLR, 2018.

[6] Xu, Mengde, et al. "End-to-end

semi-supervised object detection with soft teacher." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
[7] Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
[8] https://www.cisp.or.kr/wp-content/uploads/2019/08/빅데이터_비식별화_기술_활용_안내서_ver_1.0.pdf
[9] Akyon, Fatih Cagatay, Sinan Onur Altinuc, and Alptekin Temizel. "Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection." arXiv preprint arXiv:2202.06934 (2022).
[10] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[11] KUANG, Zhenzhong, et al. Effective De-identification Generative Adversarial Network for Face Anonymization. In: Proceedings of the 29th ACM International Conference on Multimedia. 2021. p. 3182-3191.

저 자 소 개

김현수(Hyun-Soo Kim)

2019년 단국대학교 소프트웨어학과 학사
2019년 - 현재 엘에스웨어(주) 선임
2021년 - 현재 숭실대학교
AI,SW융합학과 석사
<주관심분야> 소프트웨어 공학, 인공지능, 컴퓨터 비전, 머신러닝, 빅 데이터

이예슬(Ye-Seul Lee)

2018년 숭실대학교
융합소프트웨어학과 석사
2018년 - 현재 엘에스웨어(주) 책임
<주관심분야> 머신 러닝, 딥 러닝, 빅 데이터, 보안

신동명(Dong-Myung Shin)

2003년 대전대학교 컴퓨터공학과 박사
2001년 - 2006년 한국정보보호진흥원
응용기술팀 선임연구원
2006년 - 2014년 한국저작권위원회
저작권기술팀 팀장
2014년 - 2016년 한국스마트그리드사업단
보안인증팀 팀장
2016년 - 현재 엘에스웨어(주)
연구소장/상무이사
<주관심분야> 오픈소스 라이선스, 시스템/네트워크보안, SG인증/보안, SW취약점분석・감정, 블록체인

이찬재(Chan-jae Lee)

2017년 광운대학교 전자재료공학과 학사
2019년 - 현재 오케스트로(주) 연구원
<주관심분야> 빅데이터, 머신러닝, 인공지능, 백엔드 엔지니어링

김명호(Myung-Ho Kim)

1985년 - 1989년 숭실대학교 전산학 학사
1989년 - 1991년 포항공과대학교 분산/병렬 알고리즘 석사
1991년 - 1995년 포항공과대학교 분산/병렬 프로그램개발환경 박사
현재 숭실대학교 교수
<주관심분야> 프라이버시, 보안, 인공지능, 시스템 소프트웨어, 공개 소프트웨어