IDAP INSTITUTE OF DEFENSE ACQUISITION PROGRAM

JAMS JOURNAL OF ADVANCES IN MILITARY STUDIES

2023, Vol. 6, No. 2, 181-196.

https://doi.org/10.37944/jams.v6i2.220

Technology trends in deep learning networks for real-time object detection in drone environment

Mun, Jonghyeon* · Sohn, Chaebong**

ABSTRACT

With the recent announcement of the Defense Innovation 4.0 Basic Plan by the Ministry of National Defense, the role and operational scope of drones are expanding as a key force in AI-based, unmanned, and autonomous systems. Consequently, the significance of real-time object detection technology is emphasized as drones take on diverse missions, including delivering, analyzing, and assessing real-time, target-related information. The emergence of recent deep learning has led to substantial advancements in the field of computer vision, particularly in object detection. Deep learning-based object detection is actively being researched, with a focus on algorithms suited for embedded and mobile environments such as drones. This research predominantly aims to develop deep learning-based object detection models that ensure real-time performance and accurately identify objects' various forms and sizes. Recent object-detection models have been categorized into backbone networks, neck networks, and head networks. By utilizing these three network components, design considerations can be tailored to fulfill the requirements of drone operations. In this paper, we investigate the technology trends of deep learning network models that can be loaded into drones for real-time object detection. Thus we contribute to strengthening effective drone operation in military operations and supporting research and decision-making processes.

Keywords : drones, computer vision, deep learning network, real-time object detection

* (First Author) Kwangwoon University, Department of Electronic Communication Engineering, Master's course student, mjh110311@kw.ac.kr, https://orcid.org/0000-0002-9921-2796.

** (Corresponding Author) Kwangwoon University, Department of Defense Acquisition Program Division, Professor, cbsohn@kw.ac.kr, https://orcid.org/0000-0001-9584-7930.

Ⅰ. 서론

4차 산업혁명시대에 드론 산업이 급격하게 발전하면서 최근 국방 분야도 아군의 인력 손실이 없이 저비용으로 실시간 표적 관련 정보를 전달, 분석, 판단할 수 있는 첨단 정보 수집과 무력 투사 수단으로써 드론의 운용이 증가하고 있다.[1] 또한, 높은 기동성과 접근성 갖춘 드론을 활용한 효과적이고 유연한 데이터 수집이 가능하여 컴퓨터 비전 및 원격 탐사 분야에서 드론을 이용한 연구가 활성화되면서 드론을 이용한 객체 인식 기술이 적극적으로 적용되고 있다(Wu et al., 2021). 예를 들어, 2021년 국립소방연구원은 사람이 접근하기 어려운 곳의 실종자와 발화점을 수색하고 관제하는데 드론을 이용한 객체 인식 기술을 활용하여 구조・수색 분야의 실적이 연평균 140%, 현장 모니터링으로는 연평균 66%씩 증가하였다. 2019년 해양수산부는 최근 5년 간 해양수산 분야 드론 활성화 방안으로 드론을 이용한 해양 감시망(해양 쓰레기 및 해양오염 감시, 해양 생태 모니터링, 항만시설 및 불법행위 감시 등)을 구축해나가고 있다. 그리고 국방 분야에서는 드론을 활용한 드론봇 전투체계 발전 방안[2]으로 2018년에 육군에 드론봇 전투단을 창설하여 표적 획득 및 감시 정찰 임무에서 드론을 이용한 객체 인식 기술이 핵심적으로 활용되고 있다.

이러한 객체 탐지 기술은 컴퓨터 비전에서 가장 중요하고 핵심적인 분야 중 하나이다(Liu et al., 2020). 특히, 최근 디지털 이미지나 영상으로부터 유용한 특징을 효과적으로 추출할 수 있도록 학습하는 방식을 취하는 합성곱 신경망을 이용한 딥러닝 기술의 등장으로 발전 추세에 있다. 또한 큰 규모의 다양한 벤치마크 데이터 세트와 GPU를 이용한 향상된 컴퓨팅 성능을 기반으로 뛰어난 성능을 갖는 딥러닝 기반의 객체 탐지 기술들이 제안되고 있다(Pal, Pramanik, Maiti, & Mitra, 2021). 하지만 일반적으로 제한된 시점만을 갖는 영상과는 달리, 드론으로 촬영된 항공 영상에는 작은 크기의 객체가 불균일하게 다수 존재하며 드론에 탑재된 카메라의 움직임에 의해 객체가 급격하게 이동 및 회전하게 된다. 이외에도 모터의 진동에 의한 흔들림, 주변 환경에 따른 조도 변화 및 폐색 등으로 인해 드론으로 촬영된 항공 영상에서의 객체 탐지 기술은 매우 도전적이고 복잡하다(Du et al., 2018). 또한 드론의 비행시간, 배터리 소모 등의 비행 성능에 미치는 영향을 최소화해야 하므로 제한된 컴퓨팅 자원을 통해 빠른 처리 속도를 보장함과 동시에 객체 인식 성능을 최대화할 수 있는 객체 탐지 모델이 요구된다(Lee, Kim, Kim, & Kwon, 2021). 이에 본 논문은 이러한 임베디드 및 모바일 환경에서 드론의 비행 성능에 미치는 영향을 최소화할 뿐만 아니라, 빠른 탐지 속도 및 높은 정확도를 달성하기 위한 딥러닝 기반의 객체 탐지 모델을 3가지 네트워크 구성 요소(백본, 넥, 헤드)로 나누어 드론에 탑재할 수 있는 실시간 객체 탐지 기술 동향을 살펴보고자 한다.

---

1) 국방기술진흥연구소(2022.03.24.). 미래국방 2030 기술전략 : 드론(DRONE).

2) 한국드론혁신협회(KDIA)(2021.02.16.). 드론봇 전투체계발전방안연구.

Ⅱ. 딥러닝 기반 객체 탐지 기술 동향

2.1 딥러닝 기반 객체 탐지 모델

객체 탐지 기술은 영상 내에 존재하는 객체의 위치를 추정하고 해당 객체의 클래스를 예측하는 작업을 의미한다.[3] 기존의 객체 검출 연구는 저수준(low-level) 특징에 기반한 SIFT(Scale-Invariant Feature Transform), SURF(Speeded-Up Robust Feature), HOG(Histogram of Oriented Gradient)와 같은 방법을 활용하였다. 하지만 이러한 접근 방식의 성능 향상은 한계에 도달하였고, 딥러닝 기술의 등장과 발전을 통해 합성곱 신경망을 활용한 객체 탐지 연구가 활발히 이루어졌다. 특히 이미지넷으로 알려진 객체 탐지를 위한 이미지 데이터셋을 중심으로 진행되는 ILSVRC 대회에서 2012년 이후 딥러닝을 이용한 객체 인식 방법이 기존 방식의 성능을 뛰어넘으며, 딥러닝 기반 객체 탐지 모델이 학계에서 널리 채택되고 있다.[3]

딥러닝 기반의 객체 탐지는 객체의 위치 정보를 파악하는 영역 제안(Region Proposal)과 객체가 어떤 카테고리에 속할 것인지 대한 분류(Classification)를 수행하며, 이를 동시에 진행하면 One-stage 모델, 순차적으로 진행하면 Two-stage 모델로 구분한다. 최근 딥러닝 기반의 객체 탐지 모델의 구조는 백본 네트워크(Backbone Network), 넥 네트워크(Neck Network), 헤드 네트워크(Head Network)로 구성된다. 각 구성 요소의 조합에 따라 다양한 객체 탐지 모델이 제시되고 있다(Guo et al., 2020).

2.1.1 백본 네트워크

백본 네트워크는 객체 탐지에서 높은 탐지 정확도와 연산 효율성을 위해 입력 이미지로부터 높은 수준의 특징을 추출하는 역할을 한다(Lee, Kim, Kim, & Kwon, 2021). 대표적으로 ImageNet 데이터셋을 통해 사전에 학습된 VGGNet, ResNet, ResNeXt 등이 있으며, 이후 백본 네트워크의 높은 정확도를 얻기 위해 더 깊고 복잡한 모델들이 제안되었다(Bochkovskiy, Wang, & Liao, 2020). 하지만 드론과 같이 제한된 자원을 갖는 임베디드 장치나 모바일 장치는 높은 탐지 정확도와 작은 모델 사이즈 조건이 요구되고 있다. 그래서 최근 경량 딥러닝 기술을 활용하여 기존 백본 네트워크보다 정확도, 속도, 메모리 효율성 측면에서 좋은 성능을 갖는 경량 딥러닝 모델(DenseNet, SqueezeNet, MobileNet, EfficientNet 등)이 제안되고 있다.[4] 관련 연구동향은 다음 Table 1과 같다.

---

3) 이승재, 이근동, 이수웅, 고종국, 유원영(2018). 딥러닝 기반 객체 분류 및 검출 기술 분석 및 동향. 전자통신동향분석, 33(4), pp. 33-42.

4) 김은희, 이경하, 성원경(2020). 딥 러닝 모델의 경량화 기술 동향. 정보과학회지, 38(8), pp. 18-29.

<Table 1> Summary of backbone network

Traditional Deep Learning Network	
VGGNet (Simonyan & Zisserman, 2014)	• Improved accuracy through the progressive construction of convolutional layers and max-pooling layers. • Increased computational and memory requirements with the depth of the network.
ResNet (He, Zhang, Ren, & Sun, 2016)	• Introduces the concept of residual learning to extend the critical depth of neural network architectures. • Reduced generalization performance on small-scale or complex datasets.
ResNeXt (Xie et al., 2017)	• Introduces the concept of cardinality to create diverse groups within each layer. • Expands network width by connecting branch paths within each group in parallel. • Enhances accuracy and hardware efficiency for complex datasets.
Lightweight Deep Learning Network	
SqueezeNet (Iandola et al., 2016)	• Reduces model parameters and computational load by replacing the conventional 3x3 filters with 1x1 convolution filters. • Effective feature map extraction through the Fire module. • Optimized network architecture for embedded/mobile environments.
MobileNet (Howard et al., 2017)	• Depthwise Separable Convolution, which separates the correlation between spatial and channel dimensions. • Optimizes model accuracy and inference speed based on design requirements using only two hyper-parameters. • Maintains accuracy while providing a small model size and fast inference speed.
EfficientNet (Tan & Le, 2019)	• Proposes a Compound Scaling Method to efficiently expand network depth, width, and resolution. • High threshold performance where accuracy and efficiency gradually increase in balance as the model size expands.

2.1.1.1 VGGNet

VGG(Visual Geometry Group)에서 개발한 VGGNet은 2014년 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서 두 번째로 높은 성능을 기록한 모델로 가중치 레이어 수가 11~19개로 각각 모델 A~E를 제안하였다. 이는 깊이가 깊은 합성곱 신경망 모델을 기반으로 복잡한 패턴과 계층적 특징을 학습할 수 있는 장점이 있으나 모델의 깊이와 복잡성 때문에 모델의 요구량이 많고 학습 및 추론 속도가 느리다는 단점을 갖는다.

2.1.1.2 ResNet

ResNet은 ILSVRC 2015 대회에서 우승하면서 많이 알려지게 되었다. 중요한 아이디어는 잔여 학습(Residual Learning)을 지원하여 딥러닝의 깊은 네트워크에서 고질적으로 발생하는 과적합

(Overfitting)과 그레디언트 소실(Vanishing Gradient)을 해결함으로써 성능을 개선하였다. 잔여 학습은 입력에서 출력으로 바로 연결하는 단순 연결(Identity Connection)을 사용함으로써 깊은 네트워크에서 학습에 도움을 주는 구조이다. 기본 구조는 여러 잔여 블록(Residual Block)을 순차적으로 쌓도록 배치하여 VGG-16보다 ResNet-152가 더 적은 연산량을 보인다.

2.1.1.3 ResNeXt

ResNeXt는 ILSVRC 2016 대회에서 2등을 차지한 모델이다. 기본 구조는 ResNet과 유사하지만, 각 잔여 블록에서 그룹 컨벌루션(Group Convolution)을 사용하여 다중 분기로 학습할 수 있도록 설계되었다. 이것을 통하여 다양한 특성을 동시에 학습할 수 있고 복잡한 데이터에서도 더 좋은 성능을 발휘할 수 있는 장점이 있다. 이러한 전략은 네트워크 모델의 깊이와 너비를 증가시키는 것보다 그룹의 크기인 카디널리티(Cardinality)를 증가시키는 것이 동일한 파라미터 수를 가질 때 더 좋은 성능이 나타났다.

■ SqueezeNet

SqueezeNet은 고성능의 기기가 갖추어진 환경에서 높은 성능을 보이는 기존의 복잡한 모델들과 달리 자동차, 드론, 스마트폰과 같은 컴퓨팅 파워와 메모리가 부족한 환경에서 최적의 모델 크기와 성능을 발휘할 수 있도록 설계된 모델이다. 일반적인 합성곱 신경망 기반 모델에서 주로 사용되는 3x3 필터를 1x1 필터로 대체하여 모델 파라미터 수를 감소시키며, 1x1 필터를 이용하여 채널 수를 감소시킨 뒤 1x1 필터와 3x3 필터를 통해 다시 채널을 늘리는 파이어 모듈(Fire Module) 기법을 제안하였다. 또한 모듈마다 다운샘플링을 최소한으로만 적용하여 피처맵의 크기를 줄여가며 이미지의 정보 손실 방지함으로써 정확도 향상 및 유지한다. 이러한 새로운 계층 구조와 필터 변경을 통해 모델을 경량화하여 기존의 깊고 복잡한 모델과 거의 동일 수준의 정확도를 유지한다. 또한 적은 수의 연산량을 통해 자율주행차, 드론과 같이 실시간성이 요구되는 환경에서 활용된다.

■ MobileNet

MobileNet은 SqueezeNet과 같이 컴퓨팅 자원이 제한된 임베디드 및 모바일 환경에 적합하도록 모델 경량화 및 성능 유지에 더불어 추론 속도를 최적화하는 데 초점을 맞춘 경량화 모델이다. 기존의 합성곱 신경망 필터와 달리 공간과 채널 사이의 상관관계를 분리하여 모델 파라미터를 효율적으로 활용한 Depthwise Separable Convolution 기법을 적용하였다. 피처맵의 채널 수를 줄이는 Width Multiplier(α)와 해상도를 줄이는 Resolution Multiplier(ρ), 이 두 개의 Hyper-parameter를 활용하여 주어진 상황에 따라 모델의 추론 속도와 정확도를 조절한다. MobileNet은 주로 ImageNet 데이터셋에 대한 분류 성능을 평가지표로 하는 기존의 모델과 달리, 다양한 데이터와 도메인에서도 좋은 성능을 보여주었다. 특히 드론의 경우 배터리를 사용하는 경우가 많으므로, 이처럼 추론 속도

를 고려하여 성능을 높이는 것은 전력 효율을 향상할 수 있다.

■ EfficientNet

EfficientNet은 제한된 컴퓨팅 성능과 메모리를 고려하여 최대의 효율을 발휘할 수 있는 복합 스케일링 방법(Compound Scaling Method)을 제안하였다. 모델의 깊이, 너비, 해상도를 스케일링 요소(Scaling Factor)로 이용하여 고정된 복합 스케일링 계수(Compound Scaling Coefficient)에 따라 균형 있게 모델의 크기를 조정함으로써 모델의 정확도 및 효율성을 향상시켰다. 이러한 기존의 경화량 기술과 비교하여 복합 스케일링 방법에 따라 모델의 크기를 증가하였을 때, 모델의 정확도와 효율성이 지속적으로 증가하는 높은 임계성능을 보였다.

2.1.2 넥 네트워크

기존의 백본 네트워크만으로 항공 영상과 같이 작은 물체를 포함하는 이미지에서 객체를 정확히 탐지하는데 한계를 보인다(Ali et al., 2019). 이러한 한계를 극복하기 위해 백본 네트워크로부터 이미지나 피쳐맵의 크기를 다양한 형태로 재구성 및 재조정하여 헤드 네트워크로 전달하는 접근 방식인 넥 네트워크가 제안되었다. 이러한 넥 네트워크를 통해 이미지 내 존재하는 다양한 크기의 객체를 인식하기 위한 딥러닝 기반 객체 탐지 모델의 구성 요소로써 쓰인다(Zhu et al., 2022). 하지만 넥 네트워크를 통해 객체 인식 성능을 높일 수 있는 반면에, 모델의 추론 속도가 너무 느려질 뿐만 아니라 메모리를 많이 차지하게 된다. 하지만 이러한 문제점을 극복하고 드론과 같은 임베디드 및 모바일 환경에서 다양한 크기의 객체 인식 성능을 위해 FPN(Feature Pyramid Networks), PAN(Path Aggregation Network), Bi-FPN(Bi-directional Feature Pyramid Network) 등의 넥 네트워크가 활용되고 있다(Table 2).

■ FPN

FPN은 기존의 객체 인식을 위한 네트워크에 결합할 수 있으며, 컴퓨팅 자원을 적게 차지하면서 다양한 크기의 객체를 인식하는 방법을 제안하였다. 합성곱 신경망의 각 계층에서 생성되는 서로 다른 해상도의 피쳐맵을 쌓아 올린 형태를 갖는다. 백본 네트워크에 입력된 이미지는 순방향 경로를 통해 각 계층에서 생성된 서로 다른 해상도를 갖는 피쳐맵을 하나의 스테이지로 하는 피라미드 형태를 갖도록 한다. 최종적으로 생성된 피쳐맵을 다시 피처맵 피라미드의 각 스테이지의 크기에 맞추어 업샘플링을 순차적으로 진행한다. 그런 다음, 각 스테이지에서 생성된 피쳐맵은 1x1 필터를 통해 상위 피쳐맵의 채널 수에 따라 재조정되어 결합된다. 이와 같은 상위 고수준 피처맵을 하위 저수준 피처맵과 순차적으로 측면 결합하여 연산 경로를 하향식 경로를 생성한다. 이러한 하향식 경로를 통해 서로 다른 해상도를 갖는 피쳐맵을 결합함으로써 다양한 크기를 갖는 객체에 대한 인식 성능을 향상시켰다.

■ PAN

PAN은 백본 네트워크의 깊이가 깊어질수록 고해상도의 저수준 피쳐맵을 고수준 피쳐맵으로 전달하는 과정에서 작은 물체에 대한 정보의 손실이 발생한다. 이를 개선하기 위해 FPN의 하향식 경로를 거친 후에 다시 역방향으로 피쳐맵을 결합하는 상향식 경로를 추가한 넥 네트워크를 제안하였다. 이러한 구조는 작은 물체 탐지에 유용한 고해상도의 저수준 특징 정보의 손실을 최소화했을 뿐만 아니라, 큰 물체 탐지에서도 엣지 또는 작은 인스턴스에도 강한 특징 정보를 활용함으로써 더 정확한 객체 인식 성능을 보였다. 하지만 추가된 상향식 경로에서는 3x3 필터를 통해 각 스테이지의 크기에 따라 다운샘플링을 순차적으로 진행하기 때문에 기존 FPN 보다 많은 파라미터가 추가된다는 단점을 갖는다.

■ Bi-FPN

Bi-FPN은 PAN과 동일한 피쳐맵 피라미드 구조를 유지하며, 상향식 경로를 단순화하여 효율적인 PAN 기반 넥 네트워크를 제안하였다. PAN의 하향식 경로와 상향식 경로에서 결합된 각 피쳐맵들은 모델 성능에 대해 다른 기여도를 갖게 된다. 이 중 기여도가 적은 하나의 입력 엣지만을 갖는 결합 노드는 제거하고, 기여도에 따라 각 엣지에 가중치를 할당하여 피쳐맵을 결합하는 Weighted Feature Fusion을 제안하였다. 이처럼 입력 엣지의 기여도에 따라 네트워크를 단순화함으로써 PAN과 비교하여 더 적은 수의 파라미터와 연산량을 갖는다. 이를 통해 실시간 객체 탐지 모델의 넥 네트워크로 적용함으로써 멀티 스케일 객체 인식 성능을 개선하였다.

<Table 2> Summary of neck network

FPN (Lin et al., 2017)	• Utilizes a pyramid structure by stacking feature maps generated from the backbone network. • Creates a top-down path by resizing feature maps of smaller resolutions and combining them with lower-level feature maps. • Aids in enhancing object recognition performance for objects of various sizes.
PAN (Liu et al., 2018)	• Bidirectional information fusion that adds a bottom-up path after the top-down path of FPN. • Minimize information loss for small objects as the network depth increases. • Increases model size and computational overhead due to the addition of the bottom-up path.
Bi-FPN (Tan, Pang, & Le, 2020)	• Removal of fusion nodes based on their contribution in the PAN-based neck network structure. • Proposes Weighted Feature Fusion, which combines feature maps by allocating weights according to their contributions. • Improves real-time performance and multi-scale object recognition performance.

2.1.3 헤드 네트워크

딥러닝 기반의 객체 탐지는 객체의 위치 정보를 파악하는 후보 영역 제안과 객체가 어떤 카테고리에 속할 것인지 대한 분류를 수행하며, 이를 동시에 진행하는지 아니면 순차적으로 진행되는지에 따라 크게 One-stage와 Two-stage 모델로 구분된다(Table 3). 대표적인 Two-stage 모델은 입력 이미지 내의 객체의 위치를 제안하는 과정과 이를 기반으로 분류하는 과정이 순차적으로 진행되고, One-stage 모델은 이러한 두 처리 과정이 동시에 진행되는 것이다. 가장 대표적인 Two-stage 객체 탐지 기술은 R-CNN으로 빠른 연산을 위해 Fast R-CNN, Faster R-CNN 등의 기술들이 추가로 제안되었다. One-stage 객체 탐지 기술은 YOLO, SSD, CornerNet, CenterNet 등이 있으며, 비교적 낮은 정확도를 갖는 데 반해 빠른 속도의 추론을 수행할 수 있다.

<Table 3> Summary of head network

Head Network	
Two-stage Model	
R-CNN (Girshick et al., 2014)	• Sequentially, region proposal step through selective search algorithms and feature extraction step through CNNs for classification/regression. • Relatively slow inference speed compared to high accuracy. • Subsequent proposals like Fast R-CNN and Faster R-CNN aim to improve the inference speed aspect.
RefineDet (Zhang et al., 2018)	• Cascade approach of region proposal and classification/regression step. • Improvement of class imbalance issues through Anchor Refinement. • Enhancements in multi-scale object recognition performance through the Transfer Connection Block and Object Detection Module. • Utilizes the advantages of both one-stage and two-stage models.
One-stage Model	
YOLO (Redmon et al., 2016)	• Proposes a grid-based detection algorithm that divides input images into cells of the same size. • Relatively low accuracy compared to fast inference speed. • Subsequent updates include various versions of the model with improvements in accuracy, real-time performance and efficiency.
SSD (Liu et al., 2016)	• Generates bounding boxes with various sizes and aspect ratios through the Default Boxes Generation. • Improvement in multi-scale object recognition performance through the use of Multi-Scale Feature Maps. • Enhances multi-scale object recognition performance while maintaining fast inference speed.
CornerNet (Law & Deng, 2018)	• Proposes a keypoints-based object detection algorithm. • Generating bounding boxes of various sizes and shapes using only two pairs of keypoints. • Simple implementation and lightweight model configuration.

2.1.3.1 Two-stage 모델

■ R-CNN

R-CNN은 딥러닝 기반 객체 탐지 모델의 구조를 형성하는 데 큰 영향력을 끼친 모델이다. 후보 영역 각 단계에서 학습이 개별적으로 진행된다. 먼저 모든 입력 영상에 대해서 선택적 탐색(Selective Search) 알고리즘을 이용하여 후보 영역을 생성한다. 이를 기반으로 객체에 대한 위치 정보를 합성곱 신경망으로 학습시켜 영상 내 객체의 위치를 찾아낸다. 기존의 저수준 특징을 기반으로 하는 객채 탐지 알고리즘과 비교하여 뛰어난 성능을 보였지만, 입력 영상에서 객체 후보 영역을 생성하기 위해 선택적 탐색과 같은 별도의 처리가 필요하며, 이 알고리즘의 결과로 수천 개 이상의 후보 영역과 각 영역에 대한 합성곱 연산을 수행한다. 즉, 방대한 양의 합성곱 연산이 요구됨에 따라 많은 시간과 메모리가 요구된다. 이러한 문제점을 개선하기 위해 이후 제안된 Fast R-CNN(Girshick, 2015)은 기존의 R-CNN과 달리 먼저 전체 이미지에 대해 합성곱 신경망을 거쳐 피처맵을 생성한다. 그런 다음 생성된 피처맵으로부터 후보 영역을 제안함으로써 수많은 후보 영역에 대한 합성곱 연산으로 발생하는 병목을 해결하였다.

그 이후 제안된 Faster R-CNN(Ren et al., 2015)에서는 기존의 후보 영역 제안 알고리즘을 GPU 연산이 가능한 합성곱 신경망 기반의 후보 영역 제안 네트워크(Region Proposal Network, RPN)로 대체하였다. 이를 통해 다양한 크기와 형태의 경계 박스(Bounding box)인 앵커 박스(Anchor box)를 생성함으로써 더욱 정교한 후보 영역을 제안하게 된다. 이처럼 전체 프레임워크를 GPU 상에서 연산을 이루어져 객체 탐지를 위한 연산 속도가 빨라졌을 뿐만이 아니라, 모든 네트워크에서 피처맵을 공유하게 되어 End-to-End 학습이 가능한 구조를 갖는다.

■ RefineDet

RefineDet은 기존의 R-CNN 계열의 Two-stage 모델과 달리 후보 영역 제안하고 단계와 객체를 정확하게 분류하고 위치를 조정하는 단계가 캐스케이드 방식으로 진행된다. 먼저 클래스 불균형 문제의 원인이 되는 음성 앵커 박스를 필터링하고 앵커 박스의 위치와 크기를 조정하는 Anchor Refinement Module(ARM)을 통해 정제된 후보 영역을 제안한다. 그런 다음 FPN와 유사하게 작은 객체 탐지에 강인하도록 고안된 Transfer Connection Block(TCB)을 통해 정제된 피처맵을 Object Detection Module(ODM)로 전달하여 다양한 크기의 객체들에 대한 정확한 위치 추정과 분류를 동시에 수행한다. 이러한 구조로부터 앵커 박스를 사전에 ARM을 통해 정제한 뒤, ODM을 통해 객체의 위치 추정과 분류를 동시에 수행한다는 점에서 One-stage 모델과 Two-stage 모델의 이점을 모두 활용한 모델이며, 특히 One-stage 모델에서의 클래스 불균형 문제를 개선함으로써 더 정확한 객체 탐지 성능을 보였을 뿐만 아니라, 속도 측면에서도 기존의 Two-stage 모델보다 향상되었다.

2.1.3.2 One-stage 모델

■ YOLO

YOLO는 이미지 내의 객체의 위치 추정과 분류를 동시에 진행하는 One-stage 객체 탐지 모델 및 알고리즘이다. 이미지를 특히, 재난 및 안전 관리, 차량 감지 및 추적 분야와 같이 실시간성이 요구되는 활용 분야에서 두각을 나타낸다. YOLO 알고리즘의 동작 원리는 먼저 입력 이미지를 동일한 크기의 그리드(Grid)로 분할시켜준다. 각 그리드 셀에 대해 그리드 중심점을 기반으로 사전에 정의된 경계 박스의 개수를 예측하고 이를 기반으로 각 박스에 대한 신뢰도, 즉 객체가 존재할 확률과 해당 박스를 차지하는 비율을 계산한다. 이와 동시에 각 그리드 셀에 대해서 사전에 정의된 클래스에 대해 조건부 확률 연산을 통해 해당 객체의 클래스에 대한 분류가 진행된다. 이처럼 하나의 네트워크에서 효율적인 알고리즘을 통해 딥러닝 기반 객체 탐지에 대한 실시간성을 발전시켰다.

현재까지도 정확성, 실시간성, 효율성 등 측면에서 여러 차례로 모델과 알고리즘이 업데이트되고 있으며, 최근 업데이트된 버전으로는 YOLO-v8까지 발표가 되었다(Hussain, 2023). 그 중의 YOLO-v3(Redmon, J. & Farhadi, 2018)의 경우에 피처맵의 크기를 세 가지로 나누어 작은 객체에 대한 탐지 성능을 개선했으며, YOLO-v5(Hussain, 2023)의 경우에는 경량화 모델 기법을 통해 정확도를 유지한 채 모델의 크기와 속도에 대한 효율성을 증대시켰다.

■ SSD

SSD는 초기 One-stage 모델인 YOLO-v1의 정확도 측면에서의 한계점과 R-CNN 계열의 속도 측면에서의 한계점을 개선하기 위해 Multi Scale Feature Maps와 Default Boxes Generation이라는 두 가지 기법을 제안하였다. Multi Scale Feature Maps는 합성곱 신경망을 통해 순차적으로 이미지부터 객체에 대한 정보를 추출하는 네트워크의 중간 계층에서 각 합성곱 신경망을 생성되는 다양한 크기의 피처맵을 분류 계층으로 앞서 전달한다. 이러한 다양한 크기의 피처맵을 전달하기 위해 완전 연결 신경망 대신에 합성곱 신경망을 통해 각 피처맵에 대한 정보 손실을 최소화한다.

Default Boxes Generation은 Faster R-CNN의 다양한 크기 및 종횡비를 갖는 경계 박스, 즉 앵커 박스와 동일한 역할을 수행한다. 이를 통해 다양한 크기의 피처맵에 투영함으로써 다양한 크기 및 형태의 객체에 대한 후보 영역을 제안한다. 이처럼 SSD는 다양한 크기의 피처맵에 대해 객체 탐지를 위한 분류 및 영역 제안을 동시에 수행함으로써 기존의 Two-stage 모델의 속도 측면에서의 한계와 One-stage 모델의 다양한 크기와 형태에 대한 객체의 탐지 정확도 측면에서의 한계를 상호보완하여 정확도와 실시간성을 요구하는 응용 분야에서 딥러닝 기반의 객체 탐지에 대한 많은 발전을 이끌었다.

■ CornerNet

CornerNet은 후보 영역 제안을 위해 사용된 앵커 박스 대신 좌측 상단 꼭짓점과 우측 하단 꼭짓점으로 쌍을 이루는 특징점(key points)을 통해 경계 박스를 예측한다. 이러한 각 객체의 특징점을 학습시키기 위해 각 특징점 쌍에 대해 0과 1로 이루어진 히트맵을 사용한다. 이와 함께 CNN을 통해 이미지로부터 피처맵을 생성한 다음 수평 및 수직 방향으로 최대 풀링, 즉 코너 풀링(Corner Polling)을 진행한다. 이로부터 각 객체에 대한 특징점을 학습하여 다양한 크기와 형태의 객체에 대해 경계 박스를 생성할 수 있게 된다. 이러한 특징점을 기반으로 경계 박스를 생성함으로써 앵커 박스에 의해 발생하는 클래스 불균형 문제를 해결할 수 있을 뿐만 아니라 경계 박스 설계 시 고려해야 할 경계 박스의 크기, 종횡비와 같은 하이퍼 파라미터를 줄일 수 있다. 하지만 두 쌍의 특징점만으로는 객체의 외관 모서리 정보만을 활용하기 때문에 객체의 전체 정보 활용이 비교적 제한적이다.

이후 이런 한계를 개선한 CenterNet(Duan et al., 2019)은 객체의 중심점을 기준으로 객체를 탐지하는 방법을 제안하였다. 각 객체의 중심점을 예측하고, 이를 통해 물체의 위치와 크기를 파악한다. 또한, CornerNet과 달리 경계 박스의 형태에 대한 복잡한 예측이 필요하지 않기 때문에 구현이 간단하고 경량화된 모델을 구축할 수 있다. 이러한 간단하고 효율적인 구조를 통해 객체의 중심점을 통해 정확한 위치를 예측하여 작은 객체에 대해서도 높은 정확도를 보이며, 또한 다른 기존 방법들보다 속도 측면에서도 우수한 성능을 보였다.

Ⅲ. 결론 및 논의

본 논문은 드론에 탑재할 수 있는 실시간 객체 탐지를 위한 딥러닝 네트워크 모델 연구 동향을 탐색하였다. 드론과 같이 임베디드 및 모바일 환경에서의 객체 탐지 기술은 제한된 컴퓨팅 자원과 전력 효율을 고려하여 실시간성이 보장되는 경량화된 딥러닝 네트워크 모델이 중점적으로 연구되면서 활용되고 있다. 또한 다양한 고도에서 드론을 통해 촬영된 항공 영상에는 다양한 크기와 형태의 객체가 존재한다. 특히 군용 드론의 경우, 운용 목적에 따라 최적의 객체 탐지에 대한 정확도와 실시간성을 충족시켜야 한다. 이를 위해 객체 탐지 모델의 주요 구성 요소를 3가지 네트워크(백본, 넥, 헤드)로 나누어 임베디드 및 모바일 환경에서의 실시간 객체 탐지 분야에서 큰 발전을 이끈 딥러닝 네트워크 동향을 조사하였다.

백본 네트워크는 모델 파라미터 수를 효율적으로 줄이면서 입력 이미지의 핵심 정보를 담은 피처맵을 생성 및 전달하는 SqueezeNet, MobileNet, EfficientNet 등이 제안되었다. 넥 네트워크는 피쳐맵 피라미드 구조를 기반으로 멀티 스케일 객체 인식 성능 개선한 FPN, PAN, Bi-FPN 등이 제안되었으며 최근 객체 탐지 모델의 주요 구성 요소로 활용되었다. 헤드 네트워크는 객체 분류 및

후보 영역 제안 방식의 처리 흐름에 따라 Two-stage와 One-stage 모델로 연구가 진행되었다. Two-stage 모델의 경우, R-CNN 계열을 중점으로 높은 정확도를 확보하면서 추론 속도와 메모리 효율을 고려한 Fast R-CNN과 Faster R-CNN 구조를 통해 발전되었다. One-stage 모델의 경우, 객체 분류 및 후보 영역 제안 방식을 동시에 처리함으로써 실시간 객체 인식 모델의 발전을 이끌었으며 대표적으로 YOLO, SSD 모델을 중점으로 객체 인식 정확도를 향상되어왔다. 또한, CornerNet, CenterNet과 같이 객체의 특징점을 활용한 접근 방식을 통해 더욱 간소화된 구조의 네트워크가 제안되었다. 이처럼 최근 모델 경량화를 통해서 기존의 깊고 복잡한 네트워크 구조에서 객체 인식 성능을 유지한 채 모델의 구조를 간소화하는 접근 방식으로 실시간성이 확보된 효율적인 발전 동향을 보인다.

딥러닝 네트워크 기반의 객체 인식 모델은 이러한 주요 세 가지 구성 요소의 효과적인 조합을 통해 여러 민간 분야에서 연구가 진행되었으며 그 성능이 입증되었다. 국방 분야에서 이러한 드론에 탑재할 수 있는 실시간 객체 인식 기술은 적의 움직임을 실시간 모니터링하는데 활용될 수 있으며, 이를 통해 화력 지원 작전, 폭발물 탐지 및 제거 작전, 인질 구출 작전 등과 같은 긴급 상황에서 병사들은 더 나은 상황인식을 갖고 신속한 대응을 할 수 있을 것이다.

Acknowledgements

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author contributions

Conceptualization, Literature review, Resources and Data curation, Investigation and Methodology, Writing (Original Draft), Project administration and Supervision: MJ and SC.

Reference

Ali, S., Siddique, A., Ateş, H. F., & Güntürk, B. K. (2021). *Improved YOLOv4 for aerial object detection*. In 2021 29th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE. https://ieeexplore.ieee.org/abstract/document/9478027

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. https://doi.org/10.48550/arXiv.2004.10934

Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., ... & Tian, Q. (2018). *The unmanned aerial vehicle benchmark: Object detection and tracking*. In Proceedings of the European conference on computer vision (ECCV) (pp. 370-386). Retrieved from https://openaccess.thecvf.com/content_ECCV_2018/html/Dawei_Du_The_Unmanned_Aerial_ECCV_2018_paper.html

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). *Centernet: Keypoint triplets for object detection*. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6569-6578). https://openaccess.thecvf.com/content_ICCV_2019/html/Duan_CenterNet_Keypoint_Triplets_for_Object_Detection_ICCV_2019_paper.html

Girshick, R. (2015). *Fast r-cnn*. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448). https://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587). https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html

Guo, J., Han, K., Wang, Y., Zhang, C., Yang, Z., Wu, H., ... & Xu, C. (2020). Hit-detector: Hierarchical trinity architecture search for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11405-11414). https://openaccess.thecvf.com/content_CVPR_2020/html/Guo_HitDetector_Hierarchical_Trinity_Architecture_Search_for_Object_Detection_CVPR_2020_paper.html

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H.

(2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. https://doi.org/10.48550/arXiv.1704.04861

Hussain, M. (2023). YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines, 11*(7), 677. https://doi.org/10.3390/machines11070677

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. https://doi.org/10.48550/arXiv.1602.07360

Law, H., & Deng, J. (2018). *Cornernet: Detecting objects as paired keypoints*. In Proceedings of the European conference on computer vision (ECCV) (pp. 734-750). https://openaccess. thecvf. com/content_ECCV_2018/html/Hei_Law_CornerNet_Detecting_Objects_ECCV_2018_paper.html

Lee, J. W., Kim, J. Y., Kim, J. K., & Kwon, C. H. (2021). A Study on Realtime Drone Object Detection Using On-board Deep Learning. *Journal of the Korean Society for Aeronautical & Space Sciences, 49*(10), 883-892. https://doi.org/10.5139/JKSAS.2021.49.10.883

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature pyramid networks for object detection*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125). https://openaccess.thecvf.com/content_cvpr_2017/html/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.html

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International Journal of Computer Vision, 128*, 261-318. https://doi.org/10.1007/s11263-019-01247-4

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). *Path aggregation network for instance segmentation*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8759-8768). https://openaccess.thecvf.com/content_cvpr_2018/html/Liu_Path_Aggregation_Network_CVPR_2018_paper.html

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). *Ssd: Single shot multibox detector*. In Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 – 14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-46448-0_2

Pal, S. K., Pramanik, A., Maiti, J., & Mitra, P. (2021). Deep learning in multi-object detection and tracking: state of the art. *Applied Intelligence, 51*, 6400-6429. https://doi.org/10.1007/s10489-021-02293-7

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. https://doi.org/10.48550/arXiv.1804.02767

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster r-cnn: Towards real-time object detection with region proposal networks*. Advances in neural information processing systems, 28. https://proceedings.neurips.cc/paper_files/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. https://doi.org/10.48550/arXiv.1409.1556

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR. https://proceedings.mlr.press/v97/tan19a.html?ref=jina-ai-gmbh.ghost.io

Tan, M., Pang, R., & Le, Q. V. (2020). *Efficientdet: Scalable and efficient object detection*. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10781-10790). https://openaccess.thecvf.com/content_CVPR_2020/html/Tan_EfficientDet_Scalable_and_Efficient_Object_Detection_CVPR_2020_paper.html

Wu, X., Li, W., Hong, D., Tao, R., & Du, Q. (2021). Deep learning for unmanned aerial vehicle-based object detection and tracking: a survey. *IEEE Geoscience and Remote Sensing Magazine, 10*(1), 91-124. https://doi.org/10.1109/MGRS.2021.3115137

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500). https:openaccess.thecvf.com/content_cvpr_2017/html/Xie_Aggregated_Residual_Transformations_CVPR_2017_paper.html

Zhang, S., Wen, L., Bian, X., Lei, Z., & Li, S. Z. (2018). *Single-shot refinement neural network for object detection*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4203-4212). https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Single-Shot_Refinement_Neural_CVPR_2018_paper.html

Zhu, L., Lee, F., Cai, J., Yu, H., & Chen, Q. (2022). An improved feature pyramid network for object detection. *Neurocomputing, 483*(28), 127-139. https://doi.org/10.1016/j.neucom.2022.02.016

원 고 접 수 일 2023년 08월 04일
원 고 수 정 일 2023년 08월 24일
게 재 확 정 일 2023년 08월 30일

IDAP 광운대학교 방위사업연구소 INSTITUTE OF DEFENSE ACQUISITION PROGRAM

JAMS JOURNAL OF ADVANCES IN MILITARY STUDIES

2023. Vol. 6, No. 2, 181-196.

https://doi.org/10.37944/jams.v6i2.220

드론 환경에서 실시간 객체 탐지를 위한 딥러닝 네트워크 기술 동향*

문종현 · 손채봉*****

국문초록

최근 국방부가 국방혁신 4.0 기본계획을 발표하면서 AI 기반 무인 · 자율체계의 핵심 전력으로 드론의 역할과 운용 범위가 확대되고 있다. 이에 드론은 실시간으로 표적 관련 정보를 전달, 분석, 평가하는 것은 물론 고도의 정보 수집 등 다양한 임무를 수행하게 되면서 실시간 객체 탐지 기술의 중요성이 강조되고 있다. 최근 딥러닝 기술의 등장으로 컴퓨터 비전 분야, 특히 물체 감지 분야에서 상당한 발전이 이루어졌다. 이에 따라 드론과 같은 임베디드 및 모바일 환경에 적합한 알고리즘을 중심으로 딥러닝 기반의 객체 감지 기술이 활발히 연구되고 있다. 이러한 연구는 실시간 성능을 보장하고 다양한 형태와 크기의 물체를 정확하게 식별하는 딥러닝 기반의 물체 감지 모델 개발을 주된 목표로 한다. 최근의 딥러닝 기반 실시간 객체 검출 모델은 백본 네트워크, 넥 네트워크, 헤드 네트워크로 분류할 수 있다. 이 세 가지 네트워크 구성 요소를 활용하여 특정 운영 요구사항에 맞춰 드론 운영 요구사항을 충족하도록 설계 고려사항을 조정할 수 있다. 본 논문에서는 실시간 객체 감지를 위해 드론에 탑재할 수 있는 딥러닝 네트워크 모델의 기술 동향을 알아보고, 이를 통해 군사 작전과 국가안보 분야에서 효과적인 드론 운용을 강화하고 연구 노력과 의사 결정 과정을 지원하는 데 기여할 것으로 기대된다.

주제어 : 드론, 컴퓨터 비전, 딥러닝 네트워크, 실시간 객체 탐지

* 이 논문은 2021년도 광운대학교 특별연구학기 지원에 의하여 연구되었음.

** (제1저자) 광운대학교, 전자통신공학과, 석사과정, mjh110311@kw.ac.kr, https://orcid.org/0000-0002-9921-2796.

*** (교신저자) 광운대학교, 방위사업학과, 교수, cbsohn@kw.ac.kr, https://orcid.org/0000-0001-9584-7930.