

# YOLO를 활용한 안티-드론 객체 탐지 시스템의 학습 이미지 수에 따른 성능 분석

이영규\*, 강진호°

## Performance Analysis by the Number of Learning Images on Anti-Drone Object Detection System with YOLO

Younggyu Lee\*, Jinho Kang°

### 요 약

최근 드론으로부터 다중이용시설이나 국가 중요 시설을 보호할 수 있도록 기계 학습 기반의 실시간 안티-드론 객체 탐지 시스템이 주목받고 있다. 본 논문에서는 테러 및 전시 상황, 급작스러운 드론/무인기 침투 상황 등의 학습 데이터를 확보하기 어려운 실제 환경에서 적용할 수 있는 가이드라인을 제시하기 위해, 전이 학습기반의 YOLO 네트워크를 활용하여 학습 이미지 수에 따른 안티-드론 객체 탐지 시스템의 성능과 관계성을 분석한다.

**Key Words** : Drone, Object Detection, Machine Learning, YOLO, Number of Images

### ABSTRACT

Recently, machine learning based real-time anti-drone object detection systems have attracted great attention to protect multi-use facilities or national important facilities from drones. This paper studies the performance and relationship analysis on the anti-drone object detection system by the number of learning images with YOLO network based on

transfer learning, in order to provide guidelines that can be applied in real environments where learning data is difficult to obtain, such as terrorist and wartime situations, and sudden drone/UAV infiltration situations, and so on.

### I. 서 론

4차 산업혁명으로 과학기술 급격히 발전함에 따라 최근 드론 및 무인기의 초소형화, 저비용화, 저소음화가 이루어지고 있으며, 이에 따라 다양한 분야에 드론 또는 무인기 활용이 급격히 증가하고 있다<sup>1-3)</sup>. 반면, 발전된 기술력을 바탕으로 불법 감시, 불법 촬영, 기밀정보 유출, 테러 등에 악용되는 사례도 증가하고 있다. 따라서, 드론 또는 무인기로부터 다중이용시설이나 국가 중요 시설을 보호할 수 있는 실시간 안티-드론 탐지 시스템이 필요하다<sup>1-5)</sup>. 하지만, 드론과 무인기는 전파 반사도가 낮은 재질로 구성되고 크기가 작아, 레이더, RF, 적외선, 음향 등의 센서를 활용하는 기존의 탐지 기법은 탐지 성능에 한계가 있다. 또한, 다양한 센서들을 통합한 안티-드론 탐지 시스템은 규모가 크고 복잡하며 유지/보수 비용이 크다<sup>1-4)</sup>.

이러한 한계점을 극복하기 위해 최근 컴퓨터 비전 기반의 안티-드론 객체 탐지 시스템에 관한 연구가 국내외에서 큰 주목을 받고 있다<sup>1-4)</sup>. 컴퓨터 비전 방식은 유지 및 보수 비용이 상대적으로 낮으며 다양한 조건 아래에서도 효과적인 탐지가 가능하다. 또한, 새로운 유형의 드론이나 다양한 환경 조건에 맞게 쉽게 적용하거나 수정시킬 수 있다는 장점이 있다. 이러한 이유로, 최근 실시간 객체 탐지로 널리 활용되는 YOLO (You Only Look Once) 네트워크를 활용한 안티-드론 객체 탐지에 관한 연구가 활발히 이루어지고 있다<sup>1-4)</sup>. 논문 [1]에서는 YOLO 네트워크에서 드론 객체 탐지 시 범위 가용성을 증가하기 위해 프레임 분할을 도입하였다. 또한, 전이 학습<sup>2,3)</sup>과 Coordinate Attention 메커니즘 등의 정보 강화 기법<sup>4)</sup>을 기반으로 YOLO 네트워크의 드론 객

\* This work was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00214142), and in part by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2023-RS-2022-00156409) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation).

♦ First Author : Department of Electronic Engineering, Gyeongsang National University, lyg8984@gnu.ac.kr, 학생(학사과정), 학생회원

° Corresponding Author : (ORCID:0000-0003-2513-3304)School of Electronic Engineering, Gyeongsang National University, jinhokang@gnu.ac.kr, 조교수, 정회원

논문번호 : 202401-019-A-LU, Received January 21, 2024; Revised February 8, 2024; Accepted February 16, 2024

체 탐지 성능을 향상시킨 연구가 이루어졌다.

한편, 기계학습을 통해 우수한 성능의 객체 탐지 시스템을 구축하기 위해서는 양질의 학습 데이터가 요구된다. 일반적으로 데이터의 양이 증가할수록 성능을 향상시킬 수 있지만, 데이터의 확보와 레이블링 과정에서 높은 비용과 시간이 요구된다<sup>6,7)</sup>. 이러한 이유로, 분야와 목적에 따라 시스템이 요구하는 성능을 달성하기 위해 기계학습 모델 훈련에 요구되는 최소 이미지 수를 예측 및 이에 따른 성능 분석에 관한 연구가 최근 큰 주목을 받고 있다<sup>6,7)</sup>.

이와 같은 중요성에도 불구하고, 컴퓨터 비전 기반의 안티-드론 객체 탐지 시스템에서 기계학습을 위한 이미지 수와 객체 탐지 성능과의 관계성을 분석한 연구는 미비한 실정이다. 특히나, 테러 및 전시 상황, 급작스러운 드론/무인기 침투 상황 등의 실제 환경에서는 충분한 학습 데이터 확보가 어려우므로, 요구하는 객체 탐지 성능을 달성하기 위한 이미지 수에 관한 연구가 필수적이다. 본 논문에서는 전이 학습기반의 YOLO 네트워크를 활용하여 기계학습 모델 훈련 이미지 수에 따른 안티-드론 객체 탐지 시스템의 성능과 관계성을 분석하며, 실제 환경에 적용할 수 있도록 가이드라인을 제시한다.

## II. 전이 학습기반의 YOLO 네트워크를 활용한 안티-드론 객체 탐지 시스템

### 2.1 드론 객체 탐지를 위한 YOLO 네트워크

객체 탐지 (Object detection)란 이미지 내에서 다중 사물을 분류 (Classification)하고 더 나아가 해당 물체의 위치를 식별 (Localization)하는 기술로서, 경계 박스 (Bounding box)를 통해 이미지 내에서 드론의 존재 여부와 위치를 파악할 수 있게 해준다<sup>11,4)</sup>. 본 논문에서는 드론 실시간 탐지를 위해 객체 탐지를 위한 대표적인 딥러닝 알고리즘으로 YOLO 네트워크를 도입한다. YOLO 네트워크는 합성곱 신경망 (CNN)을 활용하는 1-stage detector 방식의 객체 검출 기법으로 다중 사물 분류와 위치 식별을 동시에 수행하기 때문에 두 단계를 순차적으로 수행하는 2-stage detector보다 처리 속도가 매우 빠르다. 또한, 이미지 내에서 사물 이외의 주변 배경까지 학습하여 환경에 따라 네트워크의 크기를 선택할 수 있고, 탐지 속도 대비 탐지 정확도가 높아 자율 주행 등 실시간 객체 탐지 및 응용 분야에 폭넓게 활용되고 있다<sup>11,4)</sup>. 2016년에 첫 번째 YOLO 네트워크가 공개된 후 현재까지 성능을 개선한 후속

네트워크들이 연구되고 있으며, 본 논문에서는 빠른 속도와 정확도, 안정성의 장점을 가져 최근 실시간 객체 탐지에 널리 활용되는 다섯 번째 버전인 YOLOv5를 도입한다<sup>4)</sup>. YOLOv5는 네트워크 구조의 크기에 따라 5가지의 모델 (n, s, m, l, x)로 구분되며, 본 논문에서는 실시간 드론 객체 탐지를 위해 크기가 가장 작아 낮은 계산 복잡도를 갖는 YOLOv5s를 활용한다<sup>4)</sup>.

### 2.2 전이 학습기반의 YOLO 네트워크를 활용한 안티-드론 객체 탐지 시스템

전이 학습은 사전학습 (Pre-training) 모델의 일부 또는 전체를 원하는 대상 데이터를 이용하여 목적에 맞게 새로운 모델로 재학습 시키는 기법으로, Upstream Task에서 사전 학습시킨 지식을 Downstream Task로 전이시킴으로써 적은 양의 학습 데이터만으로도 기계학습 성능을 향상시킬 수 있어 폭넓게 활용되고 있다<sup>2,3)</sup>. 본 논문에서는 적은 학습 데이터로도 드론 객체 탐지 시스템의 우수한 성능을 달성하기 위해 미세조정 (Fine-tuning) 기법을 도입한 전이 학습기반의 YOLO 네트워크를 설계한다.

본 논문에서 제안하는 전이 학습을 활용한 YOLO 네트워크 기반의 안티-드론 객체 탐지 시스템 구조는 그림 1과 같다. 전이 학습은 미리 학습된 모델을 특정 Task에 맞게 보다 작은 학습물을 적용하여 가중치를 미세조정하는 방식이며, Downstream Task의 데이터 전체를 사용하고 Downstream Task의 데이터에 맞게 모델 전체를 업데이트시킨다. Fine-Tuning 기법은 사전 훈련된 모델을 활용하여 오버피팅을 방지하며 모델을 처음부터 훈련하는 End-to-End 학습 방식에 비해 연산량이 작아 학습 시간을 단축시킬 수 있다.

전이 학습을 위한 사전 학습 수행 시 기계학습 네트워크의 task에 따라 대규모 데이터셋을 활용하여 모델을 학습시키는 것이 중요하다. 대규모 데이터 셋은 다양한 특징, 패턴, 분포 등을 가지기 때문

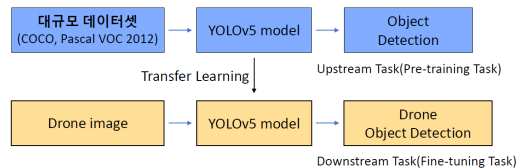


그림 1. 안티-드론 객체 탐지 시스템 구조  
Fig. 1. Structure of anti-drone object detection system

에 새로운 데이터 셋에 대한 작업을 수행할 때 특정 데이터 샘플에 대해서만 최적화되는 것을 방지함으로써 전체 네트워크의 일반화 능력을 향상시킬 수 있다. 따라서, 본 논문에서는 실시간 드론 객체 탐지를 위한 task를 고려하여 COCO 데이터셋 기반의 사전 학습 모델과 PascalVOC 2012 데이터셋 기반의 사전 학습 모델을 도입하며, 두 모델에 따른 드론 객체 탐지 성능을 분석한다. COCO 데이터셋은 주로 객체 탐지, Keypoint detection, Segmentation 등의 컴퓨터 비전 분야와 관련된 task를 목적으로 폭넓게 활용되는 표준 데이터 셋이며, 80개의 객체 카테고리를 포함하고 약 33만 장의 데이터가 존재하는 대규모 데이터셋이다. 또한, PascalVOC 2012 데이터셋은 Classification, Detection, Segmentation 등의 영상 인식 task를 위해 공개된 대규모 데이터 셋으로 20개의 객체 카테고리 및 약 10만 장의 데이터가 존재하는 대규모 데이터셋이다.

본 논문에서는, COCO 데이터셋과 PascalVOC 2012 데이터셋으로 YOLOv5s 모델을 각각 사전 학습시키고, 미세 조정 기법을 도입하여 학습된 가중치로 드론 객체 탐지 시스템을 업데이트 시킴으로써 전이 학습을 수행하여 새로운 모델의 성능을 개선한다.

### III. 이미지 수에 따른 안타-드론 객체 탐지 시스템의 성능 분석

본 논문에서는 안타-드론 객체 탐지 시스템을 위한 YOLO 네트워크의 학습을 위해 훈련 데이터 1900장, 검증 데이터 100장, 테스트 데이터 200장의 드론/UAV 이미지로 구성된 Roboflow의 Drone\_detection\_final 이미지 데이터 셋을 사용하였다. 기계학습 모델 훈련 이미지 수에 따른 안타-드론 객체 탐지 시스템의 성능과 관계성을 분석하기 위해 훈련 이미지 수를 50장에서부터 최대 훈련 이미지 수인 1900장까지 증가시키면서 학습을 진행하였다. 드론의 이동속도와 이동 방향, 장애물 등 실제 다양한 환경에 적용할 수 있도록 드론 객체 탐지 성능을 분석하기 위한 성능 지표로 mAP (Mean Average Precision), F1 Score와 mIOU(mean Intersection over Union)를 고려하였다<sup>1-4)</sup>.

그림 2는 훈련 이미지 수에 따른 mAP 및 F1 Score를 그림 3은 mIOU의 성능을 보여준다. 각 이미지 수에 따라 1900장의 이미지 중에 랜덤하게 이미지를 선택하여 학습을 10번 진행한 후 테스트 이

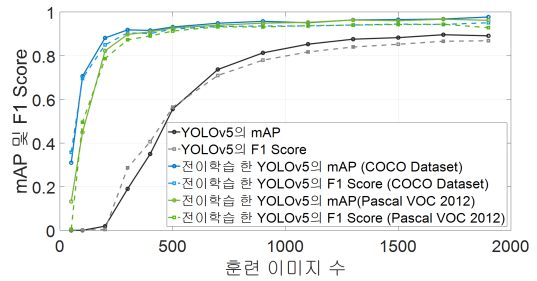


그림 2. mAP 및 F1 Score 성능  
Fig. 2. Performance of mAP and F1 Score

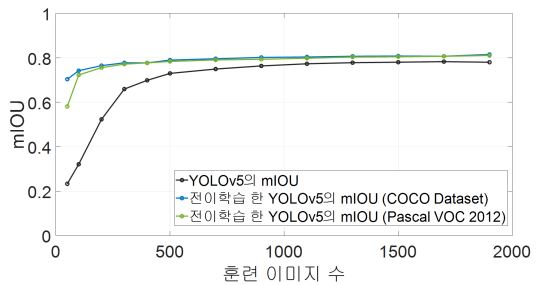


그림 3. mIOU 성능  
Fig. 3. Performance of mIOU

미지에 대한 평균 성능을 분석하였다. 전이 학습을 하지 않은 YOLO 네트워크의 경우 0.9 mAP와 0.9 F1 Score를 달성하기 위해 1900장의 훈련 이미지 수가 필요하다. 반면, COCO 데이터셋으로 전이 학습한 YOLO 네트워크의 경우 훈련 이미지 수가 300장일 때 0.92 mAP, 0.9 F1 Score, 0.77 mIOU를, 500장일 때는 0.93 mAP, 0.92 F1 Score, 0.79 mIOU를, 1900장 일 때는 0.977mAP, 0.95 F1 score, 0.815 mIOU를 달성하였다. 또한, Pascal VOC 2012 데이터셋으로 전이 학습한 YOLO 네트워크의 경우 훈련 이미지 수가 300장일 때 0.89 mAP, 0.87 F1 Score, 0.77 mIOU를, 500장일 때는 0.926 mAP, 0.91 F1 Score, 0.78 mIOU를, 1900장 일 때는 0.97mAP, 0.95 F1 score 0.81 mIOU를 달성함을 알 수 있다. 훈련 이미지 수 전 구간에 걸쳐 COCO 데이터셋으로 전이 학습한 YOLO 네트워크가 가장 우수한 객체 탐지 성능을 보인다. 이는 COCO 데이터셋이 Pascal VOC 2012보다 더 많은 객체 카테고리, 이미지를 갖는 데이터셋이므로 새로운 Task에 대한 작업을 수행할 때, 특정 샘플에 대해서만 최적화되는 것을 방지하여 전체 네트워크의 일반화 능력을 더 많이 향상 시켰기 때문이다. 한편, 훈련 이미지 수가 증가할수록 Pascal VOC

2012 데이터셋의 성능과 COCO 데이터셋의 성능 차이가 점점 감소함을 알 수 있으며, 500장 이상일 때는 두 데이터 셋의 성능이 거의 유사함을 알 수 있다.

그림 4와 그림 5는 훈련 이미지 수 대비 객체 탐지 시스템 (mAP, F1 Score, mIOU)의 성능, 즉,  $\frac{(mAP, F1Score, mIOU)}{\text{훈련 이미지 수}}$ 의 성능 지표를 보여준다.

이는 본 논문에서 새롭게 도입하는 지표로, 기계 학습에 활용되는 데이터 양 대비 객체 탐지 성능의 효율을 의미한다. 전이 학습을 하지 않은 YOLO 네트워크의 mAP와 F1 Score의 경우 이미지 수가 증가할수록 효율이 급격히 증가하다 감소하는 양상을 보이며, 500장일 때 가장 높은 효율을 달성한다. Pascal VOC 2012으로 전이학습 한 경우에는 mAP 및 F1 Score 효율이 100장까지 급격하게 증가하다가 감소하였으며, 500장 이후로는 mAP, F1 Score, mIOU 효율이 모두 완만하게 감소한다. COCO 데이터셋으로 전이학습 한 경우 이미지 수가 증가 할수록 효율이 감소한다. 한편, 이미지가 500장까지는 mAP, F1 Score, mIOU 효율이 모두 급격히 감소하는 반면, 500장 이후로는 효율이 완만하게 감소함을 알 수 있다. 이를 통해, 전이 학습기반의 YOLO

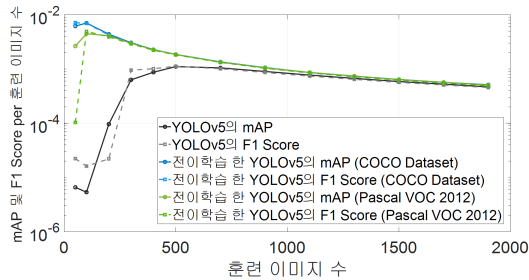


그림 4. mAP 효율 및 F1 Score 효율 성능  
Fig. 4. Performance of mAP efficiency and F1 Score efficiency

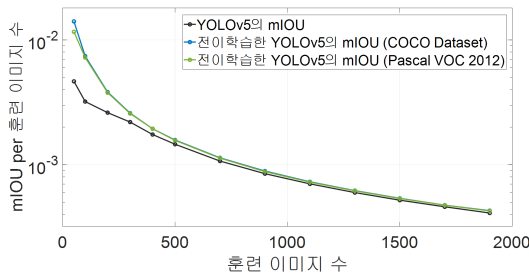


그림 5. mIOU 효율 성능  
Fig. 5. Performance of mIOU efficiency



그림 6. 안티-드론 객체 탐지 예시 (위쪽: COCO 데이터셋, 아래쪽: Pascal VOC 2012 데이터셋)  
Fig. 6. Examples of anti-drone object detection

네트워크는 학습 이미지가 500장 이상의 경우 드론 객체 탐지 성능 효율이 미비함을 알 수 있다.

그림 6은 훈련 이미지 수가 각각 100장, 300장, 500장일 때 (왼쪽, 가운데, 오른쪽) 안티-드론 객체 탐지 시스템의 예시이다. 위쪽 결과는 COCO 데이터셋으로 전이학습 한 YOLO 네트워크 결과이며, 아래쪽 결과는 Pascal VOC 2012 데이터셋 결과이다. COCO 데이터셋의 경우 Confidence Score는 100장일 때 0.43, 300장 일 때 0.75, 500장일 때는 0.77이다. 한편, Pascal VOC 2012 데이터 셋의 경우 100장일 때는 드론을 탐지하지 못하였고, 300장 일 때는 0.66으로 COCO 데이터셋의 경우보다 낮은 성능을 보인다. 반면, 500장일 때는 0.74를 달성함으로써 COCO 데이터 셋과 유사한 Confidence Score를 보인다.

따라서, YOLO 네트워크에서 전이학습을 활용할 시, COCO 데이터셋의 경우 300장, Pascal VOC 2012 데이터 셋의 경우 500장으로 비교적 작은 수의 학습 데이터양으로도 우수한 드론 객체 탐지 성능을 얻을 수 있음을 확인할 수 있다.

#### IV. 결론

본 논문에서는 전이 학습기반의 YOLO 네트워크를 활용하여 훈련 이미지 수에 따른 안티-드론 객체 탐지 시스템의 성능과 관계성을 분석하였다. 테러 및 전시 상황, 급작스러운 드론 또는 무인기 침투 상황 등 학습 데이터를 확보하기 어려운 환경에서, COCO 데이터셋 및 Pascal VOC 2012 데이터셋을

이용한 전이 학습 기반의 YOLO 네트워크를 활용하여 작은 수의 학습 데이터양으로도 우수한 드론 및 무인기 객체 탐지 성능을 얻을 수 있으며, 이미지에 여러 개의 드론/무인기가 포함되어 있거나 배경 또는 다른 객체가 있는 환경에서도 우수한 객체 탐지 성능을 달성함을 확인하였다.

wildlife monitoring,” *Ecological Informatics*, vol. 57, 2020.  
(<https://doi.org/10.1016/j.ecoinf.2020.101085>)

## References

- [1] D. Lee, et al., “Real time drone to drone detection system using YOLOv5,” in *Proc. KSC 2022*, pp. 1573-1575, Jeju Island, Korea, Dec. 2022.
- [2] N. Al-Qubaydhi, et al., “Detection of unauthorized unmanned aerial vehicles using YOLOv5 and transfer learning,” *Electronics*, vol. 11, p. 2669, Aug. 2022.  
(<https://doi.org/10.3390/electronics11172669>)
- [3] J. Singh, et al., “An ensemble-based IOT-enabled drones detection scheme for a safe community,” *IEEE Open J. Commun., Soc.*, vol. 4, pp. 1946-1956, 2023.  
(<https://doi.org/10.1109/OJCOMS.2023.3310003>)
- [4] Y. Zhao, et al., “TGC-YOLOv5: An enhanced YOLOv5 drone detection model based on transformer, GAM & CA attention mechanism,” *Drones*, vol. 7, no. 7 p. 446, 2023.  
(<https://doi.org/10.3390/drones7070446>)
- [5] J. Kang, J. H. Lee, and W. Choi, “Feedback rate sharing strategy for massive mimo systems with low-resolution converters,” *J. KICS*, vol. 47, no. 12, pp. 2047-2050, Dec. 2022.  
(<https://doi.org/10.7840/kics.2022.47.12.2047>)
- [6] F. E. Nowruzi, et al., “How much real data do we actually need: Analyzing object detection performance using synthetic and real data,” *arXiv preprint arXiv:1907.07061*, 2019.  
(<https://doi.org/10.48550/arXiv.1907.07061>)
- [7] S. Shahinfar, et al., “How many images do I need? Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous