

데이터셋 규모 축소 환경에서 YOLOX-Tiny 기반 경량 객체 탐지

김경렬, 김효원

충남대학교 전자공학과

kyungryul.kim@o.cnu.ac.kr, hyowon.kim@cnu.ac.kr

Lightweight Object Detection with YOLOX-Tiny Using a Reduced-Size Dataset

Kyungryul Kim and Hyowon Kim

Department of Electronics Engineering, Chungnam National Univ.

요약

본 논문은 제한된 연산 자원과 짧은 학습 시간이 요구되는 환경에서 경량 객체 탐지 모델의 실용 가능성을 분석한다. 이를 위해 YOLOX(you only look once) 기반 경량 모델인 YOLOX-Tiny를 사용하여, 학습 epoch 수와 COCO(common objects in context) 데이터셋 이미지 수를 제한한 단기간 학습 환경을 구성한다. 제한된 학습 조건에서도 객체 탐지 성능이 유지되는지 정량적으로 평가하였으며, COCO 평가 지표(AP)와 추론 속도를 함께 분석한다. 실험 결과, 소규모 데이터셋과 단기간 학습 환경에서도 YOLOX-Tiny는 안정적인 학습이 가능하였으며, 평균 추론 시간 약 6.04 ms로 초당 160 FPS 이상의 우수한 실시간 성능을 확인한다. 본 연구는 제한된 환경에서 경량 객체 탐지 모델의 성능과 효율성을 분석함으로써, 실제 산업 환경에서의 활용 가능성을 제시한다.

I. 서론

최근 딥러닝 기반 객체 검출(object detection) 기술은 자율주행, 로봇틱스, 지능형 감시 시스템 등 다양한 응용 분야에서 핵심적인 역할을 수행하고 있다[1]. 특히, YOLO(you only look once) 계열 모델은 단일 단계(single-stage) 구조와 높은 추론 속도와 정확도를 동시에 달성함으로써 실시간 처리가 요구되는 환경에서 연구되고 있다[2].

실제 산업 및 연구 현장에서는 고성능 GPU 환경이 항상 보장되지 않으며, 제한된 연산 자원에서는 모델 학습 및 추론 효율성에 대한 요구가 지속적으로 증가하고 있다[3]. 이러한 환경에서는 대규모 모델이나 장시간 학습이 현실적으로 어려우며, 경량 모델의 학습 효율과 성능 수렴 특성을 체계적으로 분석하는 연구가 필요하다.

객체 탐지 알고리즘 YOLOX는 anchor-free 구조와 개선된 학습을 통해 YOLO 계열의 성능을 한 단계 끌어올린 모델로 알려져 있으나 대부분의 기존 연구는 충분한 학습과 고성능 연산 환경을 전제로 성능을 평가하고 있다[4]. 반면, 제한된 학습 횟수 및 경량 모델을 대상으로 한 성능 분석 연구는 상대적으로 부족한 상황이다.

본 연구에서는 경량화 모델(YOLOX-Tiny)을 대상으로 단기간 학습 환경에서 성능과 추론 효율을 분석하고, COCO(Common objects in context) 2017 데이터셋 기반 정량 평가를 통해 경량 객체 검출 모델의 실용적 가능성을 검증한다. 특히, 학습 횟수 증가에 따른 성능 개선 가능성을 기댓값으로 제한된 학습 조건에서의 성능을 명확히 제시하는 것을 목표로 한다.

II. 본론

YOLOX는 80개의 class와 바운딩 박스에 대한 중심 좌표와 크기를 포함 하는 COCO 2017 데이터셋에서 포함하는 정보는 다음과 같다.

$$\mathbf{p}_v = [C_x, C_y, W, H, B_{\text{conf}}, C_{\text{conf}}]^T, \quad (1)$$

여기서, C_x, C_y 는 바운딩 박스의 중심 좌표, W, H 는 크기, $B_{\text{conf}}, C_{\text{conf}}$ 는 바운딩 박스와 class의 confidence score을 나타낸다. 모든 좌표 값은 입력 이미지 크기에 대해 정규화된 값으로 표현된다.

객체 탐지 알고리즘 YOLOX는 기존 YOLO 계열과 달리 anchor-free 구조를 채택한 객체 탐지 모델로, 복잡한 anchor 설계 과정 없이 객체의 중심 좌표와 크기를 직접 회귀한다. 특히 YOLOX-Tiny는 연산량과 메모리 사용량을 최소화 하도록 설계된 경량 모델로, 제한된 연산 자원을 가진 환경에서의 실시간 추론을 목표로 한다. YOLOX-Tiny는 backbone, neck, detection head로 표 1과 같이 구성되어 있다. 이러한 구조는 짧은 학습 시간과 제한된 데이터 조건에도 안정적인 성능을 확보할 수 있다.

YOLOX-Tiny의 학습은 객체 위치 추정과 객체 분류를 동시에 수행하는 다중 손실 함수 기반으로 이루지며, 전체 손실 함수는 다음과 같이 정의한다.

$$L_{\text{total}} = L_{\text{IoU}} + \lambda_1 L_{\text{L1}} + L_{\text{conf}} + L_{\text{cls}}, \quad (2)$$

여기서, L_{IoU} 는 바운딩 박스 위치 추정을 위한 IoU(intersection over union) 기반 손실 함수, L_{cls} 는 객체 클래스 분류 손실 함수, L_{conf} 는 객체 존재 여부에 대한 confidence 손실 함수, L_{L1} 는 바운딩 박스의 회귀 손실을 나타낸다. IoU loss는 예측된 바운딩 박스와 실제 박스간의 겹침 정도를 기반으로 위치 추정 정확도를 평가한다.

표 1. YOLOX-Tiny 구조

Backbone	CSPDarknet 기반 경량 네트워크, 입력 이미지로부터 다중 해상도의 특징 맵을 추출이 가능하다.
Neck	FPN 구조를 통해 다양한 해상도의 특징 정보를 통합하여 소형 객체와 대형 객체를 동시에 탐지하도록 한다.
Head	분류(classification)와 위치 회귀(regression)를 분리하여 학습함으로써 학습 안정성과 수렴 속도를 개선한다.

$$\text{IoU}_i = \frac{A_i^{(l)}}{A_i^{(u)}} \quad (3)$$

$$L_{\text{IoU}} = 1 - \text{IoU}^2 \quad (4)$$

$A_i^{(l)}$ 는 교집합 영역의 면적, $A_i^{(u)}$ 는 합집합 영역의 면적을 의미한다. IoU loss는 예측된 바운딩 박스와 실제 바운딩 박스 간의 겹침 정도를 기반으로 위치 추정 정확도를 평가한다.

III. 실험 결과

본 논문에서 학습 횟수(epochs)는 20/50/100회로 제한하고, COCO 2017 데이터셋의 이미지 수를 120K에서 25K로 축소하여 실험을 진행하였다. 실험 환경은 12th Gen Intel(R) Core(TM) i7-12650H(CPU), NVIDIA GeForce RTX 4050 Laptop(GPU), 32GB(RAM) 환경에서 진행하였으며, 객체 탐지 성능 평가는 AP(average percision), AP50, AP75, Loss, FPS(frame per second)를 기준으로 평가 하였으며, 현실적인 제한 환경을 가장한 실험 설계를 목표로 표 2와 같이 학습 조건을 설정하였다.

실험은 축소된 COCO 2017 데이터셋과 제한된 학습 횟수에도 불구하고 그림 1과 같이 손실 함수는 점진적으로 감소하며 안정적으로 수렴하는 경향을 보였다. 이는 YOLOX-Tiny의 구조가 단기간 학습 환경에서도 효과적인 특징을 보이는 것을 의미한다.

표 2. 성능평가를 위한 실험 학습 조건

항목	설정
모델	YOLOX-Tiny
데이터셋	COCO 2017 image 25K
epochs	20/50/100
batch size	8
해상도	416×416

표 3. YOLOX-Tiny model 학습 횟수에 따른 성능 지표

Epochs	AP	AP50	AP75	Loss	FPS
20	3.46	7.2	2.9	11.89	164
50	13.17	24.0	13.1	10.35	164
100	17.15	29.9	17.5	9.40	164

표 4. COCO 2017 데이터셋에서의 비교 실험

Model	Epochs	Images	AP
YOLOX-Tiny[5]	100	120K	25.40
Ours	100	25K	17.15

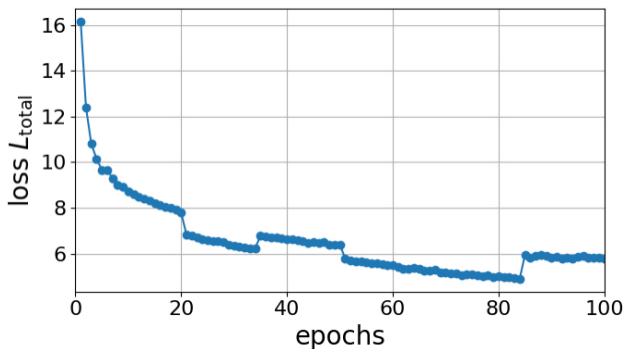


그림 1. 학습 횟수에 따른 total loss 그래프

객체 위치 추정과 클래스 예측이 정상적으로 수행됨을 확인하였고, 이는 본 연구의 목적은 최고 성능 달성이 아닌 제한된 환경에서도 YOLOX-Tiny가 실용 가능성 분석을 고려할 때 의미 있는 결과이다. 또한, 표 3와 같이 추론 성능 측면에서는 평균 6.04 ms의 추론 시간을 기록하였으며, 이는 160 FPS 이상의 실시간 처리 성능에 해당한다. 이러한 결과는 학습 자원이 제한된 환경에서도 YOLOX-Tiny가 실시간 탐지 시스템에 충분히 활용 가능함을 보여준다.

표 4에서는 COCO 데이터셋의 학습 이미지 수를 기존 120K에서 25K로 축소한 조건에서 동일하게 100 epochs를 수행하였다. 그 결과, 전체 데이터셋을 사용한 경우 AP가 25.4로 측정된 반면, 축소된 데이터셋에서는 AP가 17.15로 감소하였다. 이는 학습 epochs를 동일하게 유지하더라도 학습 데이터 규모가 성능에 직접적인 영향을 미침을 의미한다. 특히 제한된 환경에서는 모델의 일반화 성능이 현저히 저하될 수 있음을 정량적으로 확인하였다.

IV. 결론

본 논문에서는 제한된 학습 횟수와 축소된 데이터셋 환경에서 YOLO-Tiny 기반 객체 탐지 모델의 성능과 실용 가능성을 분석하였다. 실험 결과, 단기간 학습 환경에서도 안정적인 학습이 가능하며, 우수한 실시간 추론 성능을 유지함을 확인하였다.

향후 연구에서는 동일한 데이터셋 규모를 유지한 상태에서 학습 횟수를 점진적으로 증가시키며 성능과 실용 가능성을 분석할 예정이다. 또한 pose estimation과 같은 확장 과제를 적용하여 제한된 환경에서의 경량 모델의 가능성을 검증하는 연구 개발이 필요하다.

ACKNOWLEDGMENT

본 논문은 한국연구재단 기초연구사업(No. RS-2025-25415939), 한국연구재단 대학기초연구지원사업(G-LAMP)(No. RS-2025-25442707), 그리고 한국연구재단(No. NRF-2023R1A2C3002890)의 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] X. Ding *et al.*, "RepVGG: MakingVGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comp. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 13733-13742.
- [2] A. Koubaa *et al.*, "Cloud versusedge deployment strategies of real-time face recognition inference," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 143-160, Jan. 2022.
- [3] L. Liu *et al.*, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261-318, 2020.
- [4] Z. Ge *et al.*, "YOLOX: Exceeding YOLOseries in 2021," 2021, *arXiv:2107.08430*.
- [5] C. Yuxin and J. Tan, "Improved vehicle target detection algorithm based on YOLOx-tiny for lightweight remote sensing images," in *Proc. Int. Conf. Algorithms*, Aug. 2024, pp. 273-278.