

YOLO 계열 객체 탐지 신경망 모델 기술 동향

손석빈^o, 백한결^o, 박수현^o, 김중현^o

고려대학교 전기전자공학과^o

lydiasb@korea.ac.kr, 67back@korea.ac.kr, soohyun828@korea.ac.kr, joongheon@korea.ac.kr

Trends in YOLO Series Object Detection Neural Network Model Technology

Seok Bin Son^o, Hankyul Baek^o, Soohyun Park^o, Joongheon Kim^o

Dept. of Electrical and Computer Engineering Korea University^o

요약

현재 객체 탐지 기술에 대한 다양한 연구에서는 높은 성능과 실시간성을 모두 갖춘 객체 탐지 모델을 개발하기 위한 연구가 진행 중이다. 본 논문은 이러한 연구 중에서 뛰어난 성능을 보이는 YOLO (You Only Look Once) 모델에 중점을 두고 소개한다. 본 논문에서는 YOLO 모델의 특징에 대해 서술하였다. 또한, YOLO 시리즈 중에서 대표적인 모델인 scaled YOLOv4와 YOLOv7에 대해서도 자세히 소개한다.

I. 서론

현재 객체 탐지 기술에 대한 다양한 연구에서는 심층 신경망을 활용한 여러 객체 탐지 신경망 구조가 제안되고 있다 [1-3]. 이러한 구조들은 영역 제안 단계의 유무에 따라 주로 2단계 탐지 모델과 1단계 탐지 모델 분류할 수 있다. 2단계 탐지 모델의 대표적인 예로는 R-CNN 계열에 속하는 모델이 있다. R-CNN 계열에 속하는 모델들은 먼저 탐지 영역을 제안하고, 해당 영역에서 객체를 탐지하는 단계를 거친다. 이로써 정확도가 높아지는 장점이 있지만, 추가적인 연산으로 인해 추론 시간이 길어진다는 한계가 있다. 반면에 1단계 탐지 모델은 탐지 영역 제안 단계를 거치지 않고 바로 추론이 이루어지기 때문에 연산량이 적고 추론 시간이 빠르다는 특징이 있다. 따라서 실시간 성능이 요구되는 객체 검출 작업에 주로 활용된다. 이 구조의 대표적인 모델로는 YOLO (You Only Look Once) 시리즈가 있다. YOLO 시리즈는 한 번의 순전파로 모든 객체의 위치와 클래스를 동시에 예측하여 빠른 추론을 가능하게 한다. 본 논문에서는 다양한 YOLO 계열의 모델 중 대표적인 모델인 scaled YOLOv4와 YOLOv7에 대해 소개한다.

II. YOLO 모델

YOLO 모델은 객체 탐지 기술 중 대표적인 모델 중 하나이다. 이 모델은 한번의 순전파로 모든 객체의 위치와 클래스를 동시에 탐지하기 때문에 실시간 객체 탐지에 용이하다. 또한 YOLO 모델은 전역 정보를 활용하며, 이미지를 격자로 나누어 각 격자 셀이 여러 개의 Bounding Box 및 클래스를 예측하므로, 다수의 객체를 동시에 처리할 수 있다는 장점을 가지고 있다. 따라서 해당 모델은 다양한 크기와 종류를 가진 객체에도 뛰어난 탐지 성능을 보인다.

III. Scaled YOLOv4

YOLO 모델 중, Scaled YOLOv4는 객체 탐지에서 중요한 모델 스케일링 기술을 제안한 YOLOv4의 후속 모델로, 모델 크기를 다양하게 조정하여 다른 환경 및 성능 요구에 적용할 수 있다 [4]. 이 모델은 객체 탐지의

정확도를 높이면서도 추론 시간을 최소화하기 위해 깊이, 폭, 해상도, 네트워크 구조 등을 조절할 수 있는 유연한 스케일링 기법을 도입하였다. Scaled YOLOv4는 세 가지로 분류할 수 있다. 첫째, CSP-ized YOLOv4는 일반 GPU에서 실시간 객체 감지를 위해 최적화되었으며 정확도와 속도 간의 최적의 균형을 제공하여 다양한 환경에서 효율적으로 활용할 수 있다. 둘째, YOLOv4-tiny는 저사양 GPU에서도 효율적인 객체 검출이 가능하도록 설계되었다. 작은 GPU에서도 빠르고 경제적으로 객체를 검출할 수 있도록 CSPOSANet에 PCB 구조를 적용한 모델이다. 셋째, YOLOv4-Large는 CSP를 충분히 활용하여 고성능 모델인 YOLOv4-p5, YOLOv4-p6, YOLOv4-p7 등을 제공하여 클라우드 GPU와 같은 고성능 환경에서 최적의 성능을 발휘한다. Scaled YOLOv4는 YOLOv4 및 YOLOv5에 비해 레이어 수가 적은 모델에서도 메모리와 연산 간의 균형을 맞출 수 있는 장점을 가지고 있다. 또한, 다양한 규모의 모델에 적용 가능하며 최적의 속도와 정확도를 유지하면서 소규모부터 대규모까지의 네트워크에 적용할 수 있는 유연성을 제공한다.

II. YOLOv7

YOLO 모델 중, YOLOv7의 구조는 그림 1과 같으며, 다양한 하드웨어 환경에 최적화된 객체 탐지 모델로, 엣지 GPU, 일반 GPU, 클라우드 GPU를 대상으로 YOLOv7, YOLOv7-X, YOLOv7-W6이라는 세 가지 기본 모델을 제공한다 [5]. 각 모델은 성능과 파라미터 측면에서 차이가 있다. YOLOv7은 기존의 객체 탐지 방식과 다르게 'bag-of-freebies'이라 불리는 최적화된 모듈과 방법을 중점을 두어 학습 비용을 증가시키고 정확도를 향상시킨다. 모델 아키텍처 설계 시에는 파라미터 수, 계산량, 계산 밀도 등의 고려사항을 분석하여 E-ELAN (Effective and Efficient Layer-wise Adaptive Network)을 도입했다. 이는 경사 전파 경로를 유지하면서 효율적으로 계산 블록을 추가하여 모델 학습 능력을 향상시킨다. 또한, 다양한 모델 스케일링 방법과 깊이, 폭을 조정하여 다양한 성능 요구 사항을 충족시키고 RepConv라는 새로운 컨볼루션 모듈을 도입하여 모델을 재 파라미터화했다. YOLOv7은 신뢰할 수 있는 소프트 라벨을 할당하는 라벨 할당 방법을 도입하였으며, 이 중에서 auxiliary head와 lead

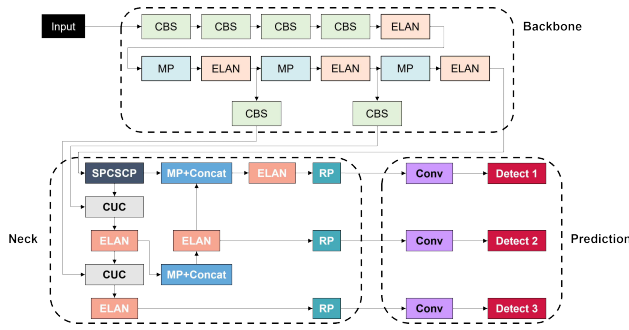


그림 1 YOLOv7 구조도

head에 소프트 레이블을 할당하는 새로운 방법을 개발했다. 더불어, 컨볼루션 레이어에 직접 일괄 정규화, 벡터화된 암묵적 지식, 지수이동평균 (Exponential Moving Average; EMA) 모델을 사용한 학습 가능한 방법을 추가로 도입하여 성능을 높였다. YOLOv7은 매개변수 수와 계산 비용을 줄이면서 평균 정확도 (Average Precision; AP)를 향상시켰으며, 작은 모델에서도 효율적인 결과를 제공하며 클라우드 GPU 모델에서도 뛰어난 성능을 보여준다.

III. 결론

현재 객체 탐지 기술에 대한 연구는 높은 성능과 동시에 실시간성을 갖춘 모델을 개발하는 방향으로 진행되고 있다. YOLO 모델은 이 분야에서 대표적인 모델로, 이미지를 단 한 번의 순전파로 처리하여 빠른 추론 시간을 제공한다. 이는 실시간 객체 탐지에 적합하며, 객체의 위치와 클래스를 한 번에 예측하는 특징이 있다. 이러한 YOLO 계열의 모델 중, 본 논문에서는 scaled YOLOv4와 YOLOv7에 중점을 두어 소개한다.

ACKNOWLEDGMENT

본 연구는 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2021-0-00766, 신경망 응용 자동생성 및 실행환경 최적화 배포를 지원하는 통합개발 프레임워크 기술개발).

참고 문헌

[1] X. He, K. Zhao, and X. Chu, "Automl: A survey of the state-of-the-art," Knowledge-Based Systems, vol. 212, pp. 106622, 2021.

[2] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, pp. 8697 - 8710, Salt Lake City, UT, USA, June 2018.

[3] Z. Zhong, J. Yan, W. Wu, J. Shao, and C.-L. Liu, "Practical block-wise neural network architecture generation," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, pp. 2423 - 2432, Salt Lake City, UT, USA, June 2018.

[4] C. Wang, A. Bochkovskiy, and H. Liao. "Scaled-yolov4: Scaling cross stage partial network." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, pp.

13029-13038, virtual, June 2021.

[5] Wang, C. Y., Bochkovskiy, A., and Liao, H. Y. M., "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, pp. 7464-7475, Vancouver, BC, Canada, June 2023.