

저성능 엣지 디바이스에서 MobileNet 기반의 보행자 탐지 및 출입 인원 계수 기법

최상일, 우혁, 조민석*
올뉴스시스템즈, *현대모비스

csi0271@allnewsystems.com, hwu@allnewsystems.com, *mixergi@mobis.co.kr

Pedestrian Detection and Entrance Counting Based on MobileNet in Low-end Edge Device

Sang Il Choi, Hyuk Wu, Min Seok Jo*
Allnewsystems, *Hyundai Mobis

요약

본 논문은 저성능 엣지 디바이스에서 실시간으로 보행자를 탐지하고 출입 인원을 계수하는 기법을 제안한다. MobileNet 모델을 기반으로 보행자 이미지 데이터셋을 이용해 전이 학습을 수행하고 적은 연산으로 다중 객체 추적이 가능한 SORT 알고리즘을 적용하여 실시간으로 출입 인원을 계수한다.

Abstract

In this paper, we present a technique for edge device for detecting and counting pedestrians entering and leaving. Transfer learning is performed using the pedestrian image dataset based on the MobileNet and it counts pedestrians in real time by applying the SORT algorithm that can track multiple objects with little computation.

I. 서론

엣지 컴퓨팅은 데이터 소스의 물리적인 위치나 그 근처에서 연산을 수행하는 것을 의미하며 이러한 연산을 수행하는 장치를 엣지 디바이스라고 한다. 최근 AI의 발전과 사물인터넷(IoT) 기기의 도입 그리고 엣지 컴퓨팅이 어우러지며 저사양의 엣지 디바이스에서도 빠른 추론 속도로 실시간 성능을 갖는 딥러닝 기법이 주목받고 있다.

딥러닝 모델은 여러 계층의 신경망으로 구성되어 있으며 라벨링된 데이터를 이용해 학습 하면서 정확한 결과를 얻을 수 있는 가중치를 업데이트 한다. 모델의 신경망이 복잡할수록 업데이트 해야하는 파라미터의 개수와 연산량이 많아지는 반면 높은 성능을 얻을 수 있다. 복잡한 신경망은 모델이 동작하는 엣지 디바이스의 높은 GPU 사양을 요구한다.

본 논문에서는 저성능의 엣지 디바이스에서 빠른 추론 속도와 정확한 성능을 얻기 위해 파라미터가 적은 MobileNet[1] 모델을 기반으로 전이 학습(Transfer Learning)[2]을 수행하고 탐지한 객체에 추적 및 할당 알고리즘을 적용하여 출입 인원을 계수하는 기법을 구현했다.

본 논문의 구성은 다음과 같다. 먼저 2 장에서는 사용한 모델에 대한 기본 개념과 전이 학습 네트워크에 대해 기술하고 모델 추론과 추론 결과를 이용한 인원 계수 기법에 대해 설명한다. 3 장에서는 논문을 마무리하며 결론을 제시한다.

II. 본론

i) 엣지 디바이스

기존 데스크탑, 랩탑, 서버 등에서 사용하는 외장 그래픽카드 기반의 모델은 그래픽 성능이 낮은 Jetson Nano 와 같은 엣지 디바이스에 사용함에 있어 성능 제약으로 인하여 사용에 어려움이 있다. 그러므로 엣지

디바이스에서는 기존 모델보다 복잡도가 낮고, 연산량이 적으면서도 높은 정확도를 만족시킬 수 있는 모델에 대한 연구가 필요하다. 이러한 요구사항을 만족하기 위하여 학습된 딥러닝 모델의 추론 속도를 수 배에서 수십 배까지 향상시켜 딥러닝 서비스를 개선하는데 도움을 줄 수 있는 모델 최적화 라이브러리인 TensorRT[3]를 사용하였다.

ii) MobileNet

1-Stage Detector[4]의 한 종류인 Single Shot Multibox Detector(SSD)[5] 모델은 입력 이미지가 6 개의 은닉층을 통과하며 각 단계마다 나온 Feature Map을 이용해 Object Detection을 수행한다.

각 단계에서 나온 탐지 결과를 이용해 Non-maximum Suppression(NMS)에서 최종 탐지 결과를 추론한다. 이를 통해 제한된 성능의 엣지 디바이스에서도 빠른 탐지 속도와 높은 성능을 확보했다.

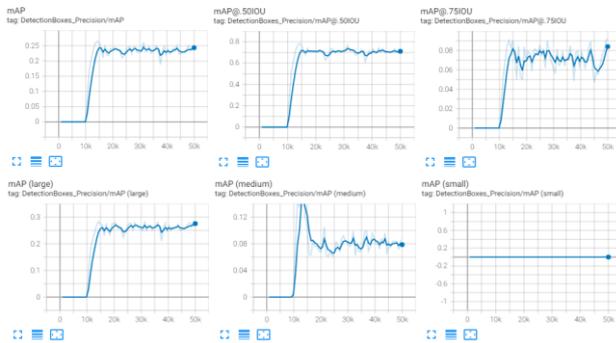
MobileNet v1 모델은 기존의 Convolution 연산 대신에 Depthwise Separable Convolution 연산을 사용한다. Depthwise Separable Convolution은 입력에 사용된 이미지를 채널별로 분리하고 각 채널에 할당한 필터와 Convolution 연산을 수행한다. 이를 통해 나온 Feature map을 병합하지 않기 때문에 Convolution만을 수행한 연산량과 비교하였을 때 입력 이미지의 채널 수만큼 연산량이 감소한다. Convolution 연산을 하는 VGG Net-D[6] 모델과 비교하여 파라미터의 양이 대폭 줄었음에도 불구하고 정확도는 약 1% 가량 차이가 났다.

iii) 모델 학습

베이스 라인 모델은 COCO[7] 데이터셋으로 사전 학습이 이루어진 모델을 사용했다. 전이 학습에 사용한 데이터 셋은 보행자를 천장에서 촬영한 이미지 5000 장을 사용했으며 학습, 검증, 테스트 데이터 비율을 7:2:1로 나누었다. 데이터 셋의 라벨은 YOLO[8]

포맷으로 작성되어 있어 포맷 변환 프로그램을 사용하여 COCO 포맷으로 변환했다. COCO 포맷으로 변환한 데이터 셋을 다시 한번 TFRecord 포맷으로 변환했다. TFRecord 는 Google 의 Protocol Buffer 라이브러리를 사용해 이미지와 라벨 데이터를 하나로 통합하여 직렬화하여 바이너리 방식으로 저장한 데이터 포맷이다. 이를 통해 학습에 사용되는 이미지와 라벨 데이터를 따로 읽어 들이는 학습 방식보다 학습 속도를 향상시켰다. 모든 입력 데이터의 크기는 300x300 사이즈로 변환하였다. 학습에 사용한 파라미터는 Batch size, Learning rate, Epoch 총 3 가지다. Batch size 는 모델이 가중치를 업데이트 하기 위한 입력 데이터의 단위로 24 로 설정했다. Learning rate 은 Loss 값이 작아지는 방향으로 가중치를 업데이트 하는 변화량으로써 0.004 로 설정했다. Epoch 는 모든 데이터를 학습하는 반복 횟수로 5000 으로 설정했다.

그림 1 은 mAP 가 0.25 정도 나타내고 mAP@.50IOU 는 0.70, mAP@.75IOU 는 0.80 의 결과를 보였다.



[그림 1] MobileNet v1 전이 학습 mAP 그래프

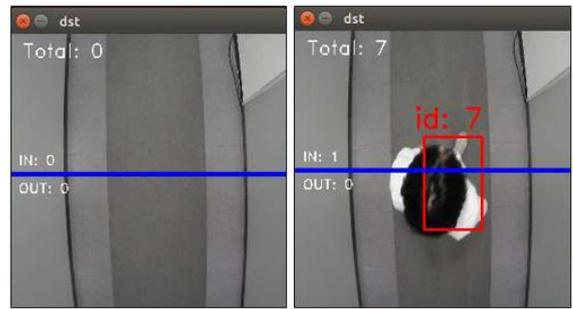
iv) 모델 추론

학습과 테스트를 완료한 모델을 추출하여 엡지 디바이스에 최적화하기 위해 TensorRT 로 변환했다. TensorRT 는 엔비디아 GPU 상에서 추론 속도를 수 배에서 수십 배 향상시킬 수 있는 모델 최적화 엔진이다. TensorRT 를 이용하여 15fps 의 속도를 보장받고 저성능 엡지 디바이스에서 보다 빠른 속도와 높은 정확도의 추론이 가능함을 확인했다.

v) 인원 계수

모델의 추론으로 탐지한 보행자의 Bounding box 를 이용해 Simple Online and Realtime Tracking(SORT)[9] 알고리즘을 적용하여 객체를 추적한다. SORT 는 다중 객체 추적이 가능하며 연산량이 적은 반면 정확도가 높은 장점을 갖고 있어 실시간 객체 추적에 적합하다.

이전 프레임에서 탐지한 보행자의 Bounding box 에 Kalman filter 를 적용하여 여러 개의 후보 bounding box 를 추정한다. 다음 프레임에서 탐지한 보행자의 bounding box 와 Intersection over Union(IoU)을 계산하여 score 가 가장 높은 박스를 동일한 보행자로 판단한다. 보행자가 여러 명인 경우 Hungarian 알고리즘을 적용하여 이전 프레임에서 탐지한 보행자에게는 동일한 식별 번호를 할당하고 최초로 탐지한 보행자에게는 새로운 번호를 할당한다. 이를 통해 출입한 총 인원을 계수하고 보행자가 가상의 선을 지나는 방향을 판단하여 입장 및 퇴장 인원을 계산한다.



[그림 2] 보행자 탐지 및 추적 결과

그림 2 는 보행자 탐지 및 추적 결과를 보여주는 실행 화면이다.

III. 결론

본 논문에서는 기존의 고성능 GPU 를 필요로 하는 보행자 탐지 모델 추론 및 출입 인원 계수를 저성능 엡지 디바이스에서 구현 가능하도록 한 기법을 제안한다. MobileNet 모델을 기반으로 보행자 이미지 데이터셋을 전이 학습한다. 학습이 완료된 모델을 TensorRT 로 변환하여 엡지 디바이스의 성능 한계를 극복했다. 또한 보행자를 추적하기 위해 SORT 알고리즘을 적용하여 가상의 선을 지나는 출입 인원을 계수했다.

ACKNOWLEDGMENT

이 논문은 2022 년도 과학기술정보통신부의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 연구임. (A1504-22-1003)

참 고 문 헌

- [1] Andrew G. H. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.
- [2] Karl W. "A survey of transfer learning," J. Big Data, vol. 3, no. 1, pp. 9, 2016.
- [3] "NVIDIA TensorRT Documentation," <https://docs.nvidia.com/deeplearning/tensorrt/developer-guide/index.html>
- [4] Zhengxia Z. "Object Detection in 20 Years: A Survey," arXiv preprint arXiv:1905.05055, 2019.
- [5] Liu W. "Ssd: Single shot multibox detector," European conference on computer vision. Springer, Cham, pp. 21-37, 2016.
- [6] Karen S. "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2015.
- [7] Lin T. Y. "Microsoft coco: Common objects in context," European conference on computer vision, Springer, Cham, pp. 740-755, 2014.
- [8] Joseph R. "You Only Look Once: Unified, Real-Time Object Detection," arXiv preprint arXiv:1506.02640, 2016.
- [9] Alex B. "Simple Online and Realtime Tracking," arXiv preprint arXiv:1602.00763. 2017.