임베디드 시스템용 Single Shot Multibox Detector Model 기반 적외선 열화상 영상의 객체검출

*나웅환 **김응태

한국산업기술대학교

*dndghks0706@naver.com**etkim@kpu.ac.kr

Object Detection of Infrared Thermal Image Based on Single Shot Multibox Detector Model for EmbeddedSystem

*NA, Woong Hwan **Kim, Eung Tae

Korea Polytechnic University

요약

지난 수 년 동안 계속해서 일반 실상 카메라를 이용한 영상분석기술에 대한 연구가 활발히 진행되고 있다. 최근에는 딥러닝 기술을 적용한 지능형 영상분석기술로 발전해 왔으며 국방기지방호, CCTV, 사용자 얼굴인식, 머신비전, 자동차, 드론 산업이 활성화되면서 많은 시너지를 효과를 일으키고 있다. 그러나 어두운 밤과 안개, 날씨, 연기 등 다양한 여건에서 따라서 카메라의 영상분석 정확성 감소와 오류가 수반될 수 있으며 일반적으로 딥러닝 기술을 활용하기 위해서는 고사양의 GPU를 필요로 하기 때문에 다른 추가적인 시스템이 요구된다. 이에 본 연구에서는 열적외선 영상의 객체 검출에 적용하기 위해 SSD(Single Shot MultiBox Detector) 기반의 경량적인 MobilNet 네트워크로 재구성하여, 모바일 기기 등 낮은 사양의 낮은 임베디드 시스템에서도 활용 할 수 있는 방법을 제안한다. 모의 실험결과 제안된 방식의 모델은 적외선 열화상 카메라에서 객체검출과 학습시간이 줄어든 것을 확인 할 수 있었다.

1. 서론

최근 영상 분석 기술 시스템이 빠르게 발전하면서 매우 다양한 분야에 파급되고 있으며, 기능 또한 단순한 주변 상황 감시형 아날로그 영상 시스템에서 자동으로 사물이나 사람의 특징적인 객체를 인식·추적할 수 있는 딥러닝 네트워크 기반의 지능형 영상 분석 시스템으로 빠르게 발전하고 있다. 더불어 일반 실상카메라에서의 영상 분석이 어려운 어두운 밤과 안개, 날씨, 연기 등 다양한 여건에서는 적외선 열화상 카메라를 활용한 영상분석 기술에 대한 연구가 진행되고 있다. 최근 국방부는 중요시설 경계시스템사업을 통해 2017년부터 2024년까지 육· 해· 공군과 국방부직할 부대의 중요시설 경계를 담당할 근거리 카메라와 철책 감지장비 등 과학화장비구축을 추진 중이며 주요시설 감시에서 악천후 및 야간 감시의 약점을 보완하는 적외선 열화상카메라 제품의 수요가 증가하고 있다. 더불어 많은 자동차 제조사들이 ADAS(AdvancedDriver AssistanceSystems)를 도입하고 있는 추세인데, 보다 높은 수준의 자율주행 기술을 구현하기 위해 객체, 다른 차량, 주변 환경시설 등을 완벽하게 인식하고 파악 할 수 있도록 적외선 열화상카메라들이 많이 활용되고 있다. 드론분야에서도 열화상카메라 탑재로 재난, 재해 현장 인명 구조에서 큰 효율을 발휘하고 있다. 이러한 여러 시스템에 딥러닝을 적용하는데 있어서 일반적으로는 클라우드 기반의 고성능 컴퓨팅 자원을 바탕으로 학습을 통해 다양한 딥러닝 모델이 만들어지지만, 기존의 학습된 모델의 정확도를 유지하면서 보다 크기가 작고, 연산을 간소화하여 모바일 기기, 경량 디바이스, IoT 디바이스 시스템에 학습된 모델을 내장하여, 지연시간 감소, 민감한 개인 정보 보호, 네트워크 트래픽 감소 같은 다양한 이점을 갖도록 하는 경량 딥러닝 모델이 요구되어지고 있다.

딥러닝을 이용한 객체를 검출하는 기법에는 구조적으로 Two-Stage 방법과 One-Stage방법이 있는데, Two-Stage방법은 입력으로 들어 온 영상에서 대략적으로 먼저 지역화(Localization)를 수행한 후 선출된 후보영역들에서 분류(Classification)와 세밀한 지역화를 하기 때문에 정확성은 매우 높으나 속도가 매우 낮은 단점이 있으며 RCNN[5],Fast R-CNN[6],Faster R-CNN[7]R-FCN[8]등이 대표적이다. One-Stage방법은 모든 영역에 대해서 지역화와 분류를 동시에 수행하여 속도가 굉장히 빠르고 Two-Stage방법에 비해 정확성은 떨어진다는 단점이 있다. 대표적으로 YOLO(You Only Look Once)[3]와 SSD(Single Shot MulitBox Detector)[1]가 있다. SSD 기법은 실시간으로 객체를 검출하는데 있어서 빠른 속도와 정확성이 가장 합리적인 성능에 이른다고 판단된다.

기존 One-Stage방법 중의 하나인 SSD(Single Shot MulitBox Detector)의 경우, 기본이 되는 네트워크 모델인 VGG-16[4]를 네트워크 구조상 층이 깊고 학습되는 파라미터의 개수가 증가하여 높은 연산량을 가짐으로 모바일 기기나 IoT환경에서는 추론속도가 매우 저하되는 문제가 발생하여 사용이 적합하지 않는 문제점이 있다. 본 연구에서는 임베디드 시스템 장

*본 연구는 중소벤처기업부 및 중소기업기술정보진흥원의 창업성장 기술개발사업의 연구결과로 수행되었음. (S-2579038)

치에 적용하여 연산량을 줄이면서도 정확도 손실을 최소화하도록 구조를 개선하기 위해서 SSD기반 MobileNet을 기본 네트워크 모델로 취하여 적외선 열화상 카메라에서의 실시간 객체검출에 대한 연구를 진행하였다.

2. 기존 관련 연구

2.1 SSD(Single Shot MultiBox Detector)

One-Stage구조의 객체 검출 기법인 SSD는 후보영역을 생성하기 위한 RPN(RegionproposalNetwork)을 따로 학습시키지 않고 특징 피라미드에서 다양한 비율의 객체에 대해서 지역화를 수행하고 모든 객체의 종류에 따른 분류를 수행한다. 즉, 다양한 크기의 특징 맵(Feature Map)을 이용하여 객체를 인식한다. CNN 기반의 기본 네트워크 모델인 VGG-16으로부터 특징 피라미드를 만드는데, 특징 맵은 합성곱 층(ConvolutionLayer)이 진행됨에 따라 크기가 줄어들게 된다. SSD는 이 과정에서 추출된 모든 특징 맵들을 추론과정에 사용하여 객체를 인식한다. 입력 영상으로부터 가까이 있는 층에서 추출되어 크기가 큰 특징 맵은 작은 물체들을 검출할 수 있고 입력 영상으로부터 멀리있는 층에서 추출되어 크기가 작은 특징 맵은 큰 물체들을 검출 할 수 있다. SSD는 RPN을 제거함으로써 Two-Stage방법들 보다 학습속도를 향상시켰으며, 다양한 크기의 특징 맵을 이용하여 보다 정확하게 객체를 인식할 수 있다.

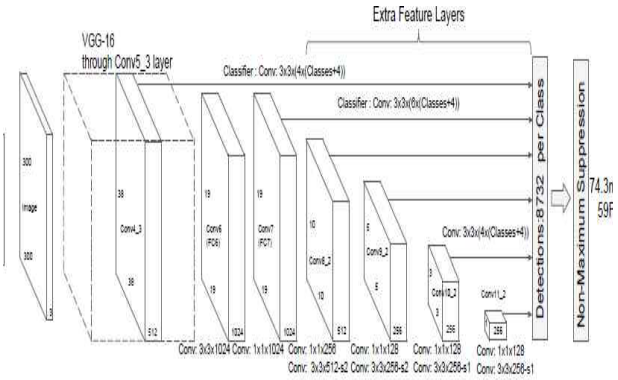


그림 1 일반적인 SSD의 시스템 구조

2.2 MobileNet

일반적인 객체 분류 및 검출 모델들은 학습 및 테스트를 위해서 많은 연산량이 요구되어 GPU 등의 고가장비를 필요로 하며, 모델 크기도 상대적으로 크기 때문에, 테스트에도 상당한 시간이 소요되었다. 특히 객체 분류 및 검출 기술을 모바일 기기, 경량 디바이스, IoT 디바이스 시스템에 활용하려는 시도를 중심으로 딥러닝 모델의 압축 및 계산량을 줄이는 고속화모델 구조로 변경하는 다양한 경량적인 딥러닝 모델들의 관한 연구가 활발히 진행되고 있다. 본 논문에서 활용한 경량적인 딥러닝 모델 중 하나인 MobileNet[2]은 기존 CNN기반 모델들에서 쓰이는 일반적인 합성곱(Convolution)은 과는 다르게 각 입력 채널마다 하나의 필터를 사용하는 깊이별 합성곱(DepthwiseConvolution)과 깊이별 합성곱의 결과를 통합하는 1x1의 위치별 합성곱(Pointwise Convolution)을 이용하여, 필터링을 담당하는 층과 통합을 담당하는 층을 분리하여 망을 설계한 DepthwiseSeparableConvolution의 형태를 구성하고 있으며, 각 각의 합성곱에는 모두 배치 정규화(BatchNormalization)와 Relu가 적용되었다. 이러한 DepthwiseSeparableConvolution의 이점으로 다른 CNN 모델들과 비교하였을 때 매우 적은 연산량과 파라미터로 높은 정확성을 도출 할 수 있다.

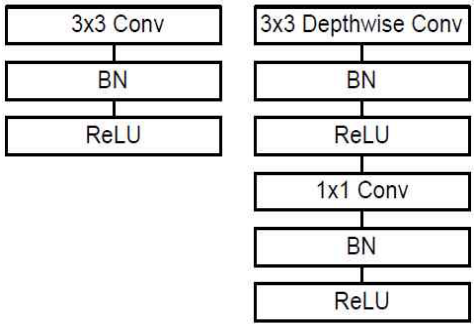


그림 2 (a) 일반 합성곱층의 구조 (b) MobileNet 합성곱층의 구조

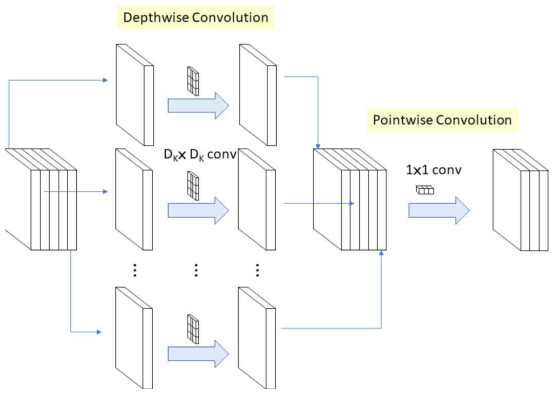


그림 3 DepthwiseSeparableConvolution의 구조

D_K를 필터사이즈, D_F입력채널사이즈, M 입력채널, N출력 채널이라 할 때 기존 합성곱에 비해 8~9배의 연산량이 감소한다.

표 1 StandardConvolution과 DepthwiseSeparableConvolution의 계산량 비교

계층	연산량
StandardConvolution	$D_K \times D_K \times M \times N \times D_F \times D_F$
DepthwiseSeparableConvolution	$D_K \times D_K \times D_F \times D_F + M \times N \times D_F \times D_F$

3. 적외선 열화상 카메라에 적용된 MobileNet-SSD 구조

기존의 SSD는 기본 CNN기반의 네트워크 모델로 VGG-16을 변형하여 활용하였다. VGG-16은 일반적인 합성곱층, 통합층(PoolingLayer), 완전 열결층(fully ConnectedLayer)으로 구성되어있는 간단한 구조를 가졌지만 네트워크의 층이 깊으며 완전 연결층이 3개가 있고 통합층을 거친 뒤에는 특징맵이 2배로 커지면서 과도하게 많은 파라미터가 존재하여 높은 연산량을 가진다. 또한 파라미터가 많다는 것은 딥러닝의 고질적인 문제인 Gradient Vanishing, Overfitting 등의 문제가 발생할 가능성이 크다. 따라서 VGGNet의 특성상 학습시키는데 오랜 시간이 걸리며, 처리속도가 비교적 느리다는 단점이 있다. 이에 본 논문에서는

저사양의 임베디드 시스템에 학습된 딥러닝 모델을 내장하여 객체 검출을 하기 위해서 SSD의 기본 네트워크 모델을 VGGNet이 아닌 MobileNet을 적용하여 연구하였으며, SSD에 이용한 MobileNet의 구조는 그림 4와 같다. 입력으로 들어온 영상은 먼저 가장 첫번째로 일반적인 3X3의 합성곱 연산을 하고 계산량과 파라미터를 대폭줄이는 MobileNet의 3X3 DepthwiseConvolution과 1X1 PointwiseConvolution으로 구성된 Depth SeparableConvoltion을 총 13층을 거쳐 SSD의 합성곱층에서 각 각의 크기가 다른 특징맵을 추출하면서 다른 객체 검출 기법들 보다 빠르고 정확성 있게 학습하고 객체를 검출 할 수 있게 된다.

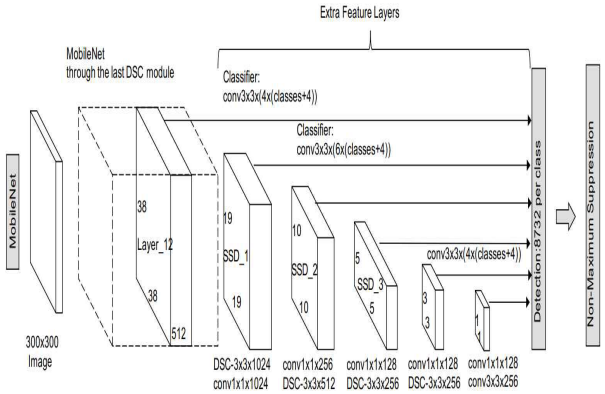


그림 4 MobileNet기반의 SSD의 시스템 구조

4. 모의 실험 결과

4.1 실험 환경

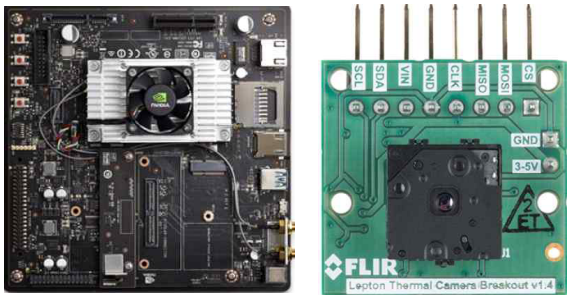
실험을 위해 임베디드 시스템 장치인 JetosnTX2와 장파 적외선 열화상 카메라 모듈인 Lepton 3.5를 활용하여 환경을 구성하였다. Jetson TX2는 딥 러닝, 컴퓨터 비전, 가속 컴퓨팅, 멀티미디어용 라이브러리들을 제공하는 전용 SDK인 Jetpack을 적하여, AI 기반 NVR(네트워크 비디오 레코더)에서 고정밀 제조 분야의 AOI(자동 광학 검사), AMR(오토노머스 모바일 로봇), 드론, 스마트 카메라 등을 구현하는데 활용되는 소형 컴퓨터이며, Lepton 3는 스마트폰 및 기타 모바일 장치에서도 사용 할 수 있는 정도의 작은 크기이며, 열 감지로 야간 방범카메라, 콘크리트 측정기, 열화상 멀티메터 등에 활용되는 카메라이다. 두 장치의 사양은 표2와 같다.

표 2 JetsonTX2의 사양

JetsonTX 2	
GPU	256-corePascal
CPU	Cortex-A57(quadcore)Denver2(dualcore)
메모리	8GB 128bit LPDDR4
저장	32GB eMMC 5.1
크기	87mm x 50 mm

표 3 Lepton 3.5 모듈의 사양

Lepton 3.5	
배열형식	160 x 120
시야각(대각선, 수평)	71°, 50°
열 감지 범위(최대)	-10°C ~450°C
유효 프레임률	8.8Hz
화소 크기	12 μm
스펙트럼 범위	장파 적외선, 8μ m ~14μ m



(a) JetsonTX2 (b) Lepton 3.5

그림 5 구현된 임베디드 보드와 열화상 센서

4.2 학습

MobileNet-SSD를 학습시키기 위해서 JetsonTX2에서 I2C와 SPI통신으로 Lepton 3.5을 연결하고 동작시켜 야간에서 사람만 촬영하였다. 열적외선 카메라에서는 출력되는 영상은 객체를 더 잘 분간 할 수 있도록 하기 위해서 온도가 높을수록 흰색 낮을수록 검은색으로 표현하였으며, 160 x 120 크기의 영상을 Bicubic 보간법을 통해 320 x 240 영상으로 크기를 2배 늘렸다. 훈련 영상과 시험 영상 각 각 4950장, 990장 총 5940장을 레이블링 한 후에 CUDA9.0v, cuDNNI7.1,1v 라이브러리와 GeForce GTX 1060 6GB의 GPU를 사용하여 학습시켰으며, 각 각의 훈련과 시험 영상들은 표4와 같이 세분화된 영상들로 구성되어 있다.

표 4 데이터 세트 구성

훈련 영상			
	사람 1명	사람 2명	사람 3명 이상
5M 거리	150 장	150장	150장
10~100M거리 (10M 간격)	거리 당 150장	거리 당 150장	거리 당 50 장
총 영상 수	4950장		

시험 영상			
	사람 1명	사람 2명	사람 3명 이상
5M 거리	30 장	30장	30장
10~100M거리 (10M 간격)	거리 당 30장	거리 당 30장	거리 당 30장
총 영상 수	990장		

5. 결과 및 결론

표5는 MobileNet기반의 SSD와 기존의 VGG-16기반의 SSD의 열화상 영상에서의 객체 검출 성능과 파라미터의 수, 학습시간을 비교 분석한 결과이다. MobileNet-SSD보다 VGG-16SSD가 mAP가 높지만, 파라미터의 수는 MobileNet-SSD가 5배, 연산량은 20배 정도 낮으며, 학습시간도 줄어들었다.

표 5 열화상 영상 데이터 세트에서의 딥러닝 모델의 비교

딥러닝 모델	mAP	파라미터의 수 (Million)	연산량 (Billion)	학습시간
VGG-16SSD	80.4%	38.2M	45	60시간
MobileNet-SSD	77.9%	7.6M	2.1	28시간

또 한 그림 6, 7, 8에서 MobileNet기반의 SSD 모델로 적외선 열화상 카메라로 거리에 따라 객체 검출을 하였으며, 최대 100m 거리에 있는 객체 까지 검출 할 수 있는 것을 확인하였다.

그림 6 MobileNet기반SSD 열적외선 카메라에서의 객체 검출(5M)

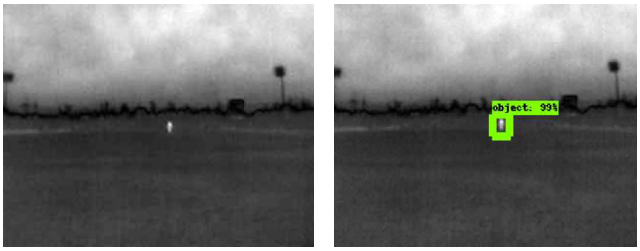


그림 7 MobileNet기반 SSD 열적외선 카메라에서의 객체 검출(50M)

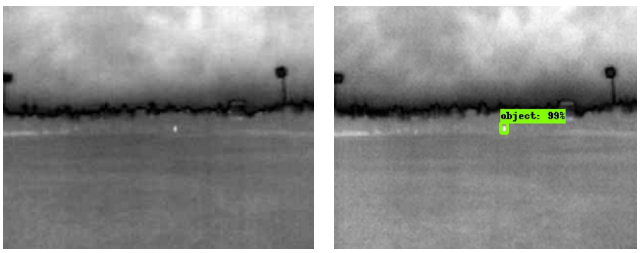


그림 8 MobileNet기반 SSD 열적외선 카메라에서의 객체 검출(100M)

참 고 문 헌

[1] W. Liu, et al.,"SSD: Single Shot MultiBox Detecto,"In ECCV, 2016.

[2] A. G. Howard, et al.,"Mobilenets: Efficient convolutional neural networks for mobile vision applications",arXiv preprint arXiv:1704.04861,2017.

[3] J. Redmon,et al.,"You Only Look Once: Unified, Real-Time Object Detection,", in CVPR, 2016.

[4] J. Redmon, et al.,"Very deep convolutional networks for large-sclaleimage recognition," in CVPR, 2016.

[5] R. Girshick, et al.,"FRegion-Based Convolutional Networks for Accurate Object Detection and Segmentation,"In TPAM, 2016.

[6] R. Girshick,"Fast r-cnn,"In ICCV, 2015.

[7] S. Ren, et al.,"Faster r-cnn: Towardsreal-timeobject detection with region proposalnetworks,"In TPAMI, 2017.

[8] J. Dai. et al.,,"R-FCN: Object Detection via Region-based Fully Convolutional Networks,"In NIPS, 2016.