

Low Resolution Infrared Image Deep Convolution Neural Network for Embedded System

Yong-hee Hong*, Sang-hun Jin*, Dae-hyeon Kim*, Ho-Jin Jhee*

*Senior Researcher, Dept. of PGM Seeker Lab, LIG Nex1, Yongin, Korea

*Chief Researcher, Dept. of PGM Seeker Lab, LIG Nex1, Yongin, Korea

*Senior Researcher, Dept. of PGM Seeker Lab, LIG Nex1, Yongin, Korea

*Chief Researcher, Dept. of PGM Seeker Lab, LIG Nex1, Yongin, Korea

[Abstract]

In this paper, we propose reinforced VGG style network structure for low performance embedded system to classify low resolution infrared image. The combination of reinforced VGG style network structure and global average pooling makes lower computational complexity and higher accuracy. The proposed method classify the synthesize image which have 9 class 3,723,328ea images made from OKTAL-SE tool. The reinforced VGG style network structure composed of 4 filters on input and 16 filters on output from max pooling layer shows about 34% lower computational complexity and about 2.4% higher accuracy then the first parameter minimized network structure made for embedded system composed of 8 filters on input and 8 filters on output from max pooling layer. Finally we get 96.1% accuracy model. Additionally we confirmed the about 31% lower inference lead time in ported C code.

▶ **Key words:** Deep Learning, CNN, VGG, Low Resolution, Infrared, Synthesize Image

[요 약]

본 논문은 저해상도 적외선영상을 사양이 낮은 임베디드 시스템에서 추론 가능하도록 강화된 VGG 스타일과 Global Average Pooling 조합으로 정확도를 증가시키면서 연산량을 최소화하는 딥러닝 컨볼루션 신경망을 이용한 저해상도 적외선 표적 분류 방법을 제안한다. 제안한 알고리즘은 OKTAL-SE로 생성한 합성영상 클래스 9개 3,723,328개를 분류하였다. 최초 임베디드 추론 가능하도록 파라미터 수가 최소화된 최대풀링 레이어 기준 입력단 8개와 출력단 8개 조합에 비해 강화된 VGG 스타일을 적용한 입력단 4개와 출력단 16개 필터수 조합을 이용하여 연산량은 약 34% 감소시켰으며, 정확도는 약 2.4% 증가시켜 최종 정확도 96.1%을 획득하였다. 추가로 C 코드로 포팅하여 수행시간을 확인하였으며, 줄어든 연산량 만큼 수행 시간이 약 32% 줄어든 것을 확인할 수 있었다.

▶ **주제어:** 딥러닝, 컨볼루션 신경망, VGG, 저해상도, 적외선, 합성영상

-
- First Author: Yong-hee Hong, Corresponding Author: Ho-Jin Jhee
 - *Yong-hee Hong (yonghee.hong@lignex1.com), Dept. of PGM Seeker Lab, LIG Nex1
 - *Sang-hun Jin (sanghunjin@lignex1.com), Dept. of PGM Seeker Lab, LIG Nex1
 - *Dae-hyeon Kim (daehyeon.kim01@lignex1.com), Dept. of PGM Seeker Lab, LIG Nex1
 - *Ho-Jin Jhee (hojin.jhee@lignex1.com), Dept. of PGM Seeker Lab, LIG Nex1
 - Received: 2021. 04. 15, Revised: 2021. 06. 01, Accepted: 2021. 06. 08.

I. Introduction

현재 의료, 군사, 과학 등 여러 분야에 적용되고 있는 적외선센서가 활발히 사용되어지고 있다. 하지만 적외선 센서는 다음과 같은 현실적인 이유로 저해상도 센서를 주로 사용할 수 밖에 없는 환경이다.

현재 가시광센서는 보통 안정적인 실리콘 베이스 공정으로 제작되어 가격도 저렴하며 성능도 좋으며 수요층도 넓어 많은 사용자를 확보하고 있으며 자연스럽게 많은 가시광영상 데이터 확보가 쉽게 가능하다. 하지만 적외선 영상은 실리콘 베이스 공정을 주로 사용하여 제조하는 가시광센서와는 달리 InSb, HgCdTe, InGaAs, VOx, TiOx 등과 같은 복합소재를 주로 이용하여 제작하기 때문에 제조공정이 복잡하다. 이러한 이유로 단소재 실리콘 제조공정에 비해 비용증가로 이어지며 이외에도 수요층이 한정적이기 때문에 가시광센서에 비해 가격이 높다. 한편 비교적 넓은 면적 초점평면배열(Focal Plane Array) 생산단가가 높기 때문에 적외선센서는 가시광센서에 비해 저해상도 센서가 주로 생산되며 수요층도 한정적이어서 확보 가능한 적외선 영상데이터도 제한적이다[1]. 이러한 이유로 가시광영상에 비해 적외선영상은 상대적으로 해상도가 낮을 수 밖에 없으며 확보 가능한 데이터도 제한적이다.

심층신경망 딥러닝은 인터넷에 의한 데이터 증가, GPU와 같은 연산가속장비 발전, CNN, ReLU, 드롭아웃, 초기화 기법 등 발전으로 기존 신경망이 가지는 문제를 대부분 해결할 수 있었으며 충분한 신경망 크기와 학습데이터가 충분하다면 종래 학습 알고리즘에 비해 더욱 높은 정확도로 학습 가능한 것이 장점이다.

가시광 영상을 활용한 딥러닝은 아래와 같이 발전해 왔다. AlexNet은 두 개 병렬처리 파이프라인을 사용하였으며 이를 GPU 분할 구조를 적용하였다. 두 개 병렬처리 파이프라인은 완전히 연결되어 있지 않아 가중치를 줄일 수 있었으며 더욱 빠르게 학습이 가능하다[2]. ZFNet은 AlexNet과 비슷하지만 첫 층 필터 크기를 AlexNet 11×11×3에서 7×7×3으로 보폭을 4에서 2로 마지막 연결 층 필터 수를 늘려 AlexNet에 비해 정확도를 높였다[3]. VGG는 필터크기를 3×3으로 줄이고 깊이를 늘려서 전체 신경망 깊이를 늘렸으며 이를 통해 비선형성과 정규화를 강화시켜 높은 정확도의 모델을 획득하였다[4]. GoogLeNet은 신경망 안의 신경망 구조 Inception 모듈을 도입하였으며 Full Connection 대신 Average Pooling을 이용하여 연산량과 가중치 수를 줄였다[5]. ResNet은 VGG를 기본으로 하였으며 Skip Connection을

도입하여 연산량을 크게 늘리지 않으면서 기울기 흐름을 개선시켜 더욱 안정적인 학습과 더욱 높은 최적화를 달성하였다[6]. DenseNet은 각 층을 feed-forward 방식으로 모두 연결하여 기존 CNN보다 적은 가중치, Vanishing-Gradient문제 완화, Regularizing 효과와 Overfitting 감소시켰다[7].

딥러닝의 높은 연산량 감소를 위해 아래와 같은 노력이 있었다. MobileNetV1은 컨볼루션 연산을 Depthwise 컨볼루션과 Pointwise 컨볼루션으로 구분하여 구현하였으며, 이로 인해 비슷한 정확도를 얻으면서 모바일 기기에서 사용 가능할 만큼 연산량을 대폭 줄일 수 있었다[8]. MobileNetV2는 기존보다 적은 채널수로 출력하는 Inverted Residual Block과 ReLU6를 이용하여, MobileNetV1보다 연산량을 줄이며 높은 정확도를 달성하였다[9]. ShuffleNet은 MobileNet의 1×1 컨볼루션 연산에 채널을 골고루 섞은 그룹 컨볼루션을 적용하여 연산량을 줄였다[10].

본 논문은 II장에서 적외선영상을 이용한 딥러닝 연구 해외 및 국내 동향을 조사하였다. III장에서 합성영상 학습 데이터 준비, 강화된 VGG 스타일 딥러닝 컨볼루션 신경망 모델 구조, 학습 환경 설정 및 과정, 학습 결과, C 코드 포팅 후 추론 수행시간 비교를 기술하였다. IV장에서는 제안하는 임베디드 시스템용 강화된 VGG 스타일 딥러닝 컨볼루션 신경망 모델 결과를 기술한다.

II. Preliminaries

1. Related works

1.1 Overseas Trend

해외 적외선 영상 딥러닝은 아래와 같이 연구되고 있다. d'Acromont는 부족한 적외선 영상 데이터를 보완하기 위해 합성영상과 Global Average Pooling을 이용하여 학습 후에 현실 영상에 적용하여 사용 가능함을 확인하였다[11]. Kim은 OKTAL-SE를 이용하여 부족한 적외선 영상 데이터를 극복하였으며, 적외선 변형 최적화(IVO)를 이용하여 비교적 일정한 영상을 획득하여 컨볼루션 신경망에 학습하는 방법을 이용하였다[12]. Park은 적외선 영상원본, 배경영상, 배경과 차이영상 3채널 영상으로 ResNet과 Atrous 컨볼루션 두 신경망을 학습하여 사람 감시정찰용 시스템을 구현하였다[13]. Akula는 적외선 영상에 CNN을 적용하여 6가지 액션 클래스 서기, 책상에 앉기, 의자에 앉기, 책상에 쓰러짐, 쓰러짐, 걷기 인식을 구현하였다[14].

1.2 Domestic Trend

국내 적외선 영상 딥러닝은 아래와 같이 연구되고 있다. 광필원은 지식기반 전이학습을 이용하여 가시광영상으로 적외선영상이 학습 가능하도록 하여 이중 센서 간 학습 방법을 제안했다[15]. 류준환은 평균 감산 필터로 표적이 존재할만한 후보 영역을 생성 후, 표적 식별용 5층 CNN 신경망을 이용하여 드론과 같은 소형 표적을 검출하였다[16]. 양성목은 기존 가시광센서로만 구성된 인체감지 감시카메라에 적외선영상과 SSD(Single Shot Multibox Detector) 딥러닝 알고리즘을 이용하여 가시광영상만 이용했을 경우에 비해 저장장치 수명을 늘렸으며, 사용용량을 줄였다[17]. 오원식은 작업자 신원 노출이 적은 적외선 영상과 CNN을 이용하여 사람이 아닌 알고리즘을 통해 인권침해를 예방하면서 작업자 부주의 실시간 모니터링을 구현하였다[18].

III. The Proposed Scheme

1. Synthesized Infrared Image

본 논문에서는 부족한 학습 및 테스트용 적외선영상 데이터를 OKTAL-SE 프로그램을 이용하여 생성하였다. OKTAL-SE는 물체의 적외선 특성 및 형상을 데이터베이스화하여 보유하고 있다. 이를 이용하여 저해상도 적외선 영상에서 분류 가능한 수준을 고려하여 그림 1과 같이 차종 분류를 적용하였다.



Fig. 1. Synthesized Infrared Image of Vehicles (From top left CW : Tank, BTR, ML, Jeep, SUV, Truck, Sedan, Pickup)

합성영상은 그림 2와 그림3과 같이 Background, Tank, BTR, ML(Missile Luncer), Jeep, Pickup, Sedan, Truck, SUV 9종을 생성하였다. 배경 이외 타겟 8종은 그림 4와 같이 타겟 중심 기준으로 진입각 범위 0~90° 간격 10°, 표적 자세각 범위 0~360° 간격 10°, 거리 범위 300~800m 간격 14m로 시야각 20°×15° 기준으로 640×480 영상을 생성하였다. 생성된 9종 적외선영상을 그림 2와 같이 정사각 영상크기에 최대한 크게 자리 잡을 수

있도록 위치시켰으며, 이를 28×28 크기로 약간 차이가 있는 두 가지 배율로 리사이즈하였다. 밝기값을 영상 내에 최소값과 최대값을 기준으로 0에서 1 값으로 노말라이즈 수행하여 신경망 입력으로 이용하였다. 이러한 방식으로 생성된 표적 영상은 28×28 크기로 리사이즈되어 있지만 실제 표적 크기가 약 10~15 픽셀 정도 영상이 대부분이다.

Background 영상 데이터는 그림 3처럼 표적 영상 생성 시와 비슷한 조건으로 OKTAL-SE 표적 외 배경 영상을 이용하여 가능한 다양한 영상을 획득하였으며 좌우 대칭 및 상하 대칭 영상을 보강데이터로 추가하여 가능한 여러 가지 경우의 Background 영상을 학습에 이용하였다.

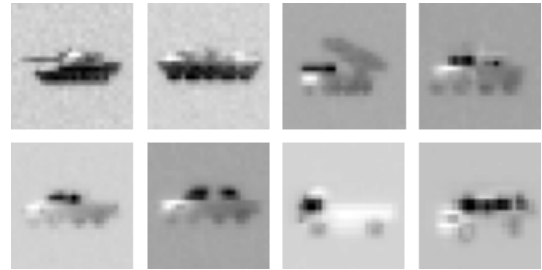


Fig. 2. Resized and Cropped Infrared Image of Vehicles (From top left CW : Tank, BTR, ML, Jeep, SUV, Truck, Sedan, Pickup)

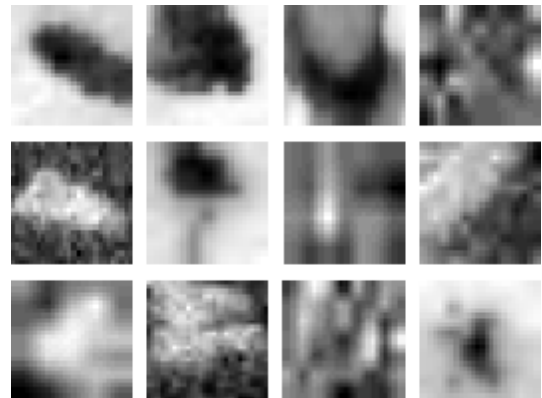


Fig. 3. Synthesized Infrared Image of Background

표 1과 같이 Background 784,000개, Tank 629,856개, BTR 209,952개, ML(2종) 419,904개, Jeep(2종) 419,904개, Pickup(3종) 629,856개, Sedan 209,952개, Truck 209,952개, SUV 209,952개를 학습에 적용하였다.

전체 학습 데이터 중 80%를 학습용 데이터로 할당하였으며 그 중 10%를 확인용 데이터로 배정하였다. 나머지 20% 데이터는 테스트용 데이터로 배정하여 학습을 진행하였다.

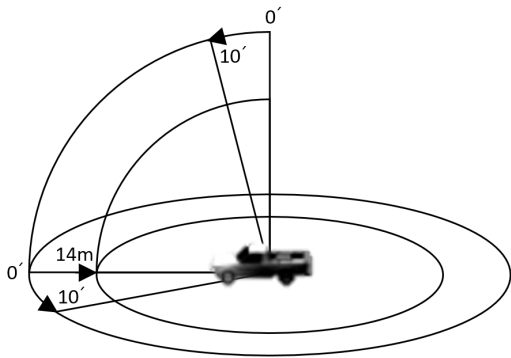


Fig. 4. Synthesized Infrared Image Generating Rule

Table 1. Synthesized Infrared Image

Class	Amount	Note
Background	784,000	Argumented
Tank	629,856	3 Types
BTR	209,952	
ML	419,904	2 Types
Jeep	419,904	2 Types
Pickup	629,856	3 Types
Sedan	209,952	
Truck	209,952	
SUV	209,952	

2. Reinforced VGG Network Structure

본 논문에서는 연산량을 줄이면서 분류 성능을 최대한 끌어내기 위해 기존 VGG 스타일 신경망을 강화하는 형태를 사용하였다. 기존 VGG 신경망은 입력단 컨볼루션 레이어에서 특정 수의 3x3 필터를 적용 후 최대풀링(Max Pooling) 레이어를 거친 후 2배씩 늘어난 3x3 필터를 적용하는 방식을 연이어 사용하였다. 하지만 시스템 성능이 많이 부족한 임베디드 시스템에서는 이러한 연산 과정은 부담된다. 본 논문에서는 기존 VGG 신경망 스타일에서 최대풀링 레이어 입력단의 컨볼루션 신경망의 필터를 기존보다 줄이고 최대풀링 레이어 출력단의 컨볼루션 신경망의 필터수를 기존보다 늘림으로서 연산량을 줄이면서도 학습 정확도를 높일 수 있었다.

여러 가지 모델을 생성 비교 실험을 위해 28x28 저해상도 영상에 적합하며 연산량을 최소화 할 수 있도록 기본 구조를 그림 5와 같이 구성하였다.

그림 5는 대표로 표 2의 모델 D를 표현하였다. 28x28 1채널 영상을 입력 받은 뒤 이를 3x3 컨볼루션 연산을 두 번 수행한 후 2x2 최대풀링을 수행한다. 이후 12x12x4로 줄어든 영상을 다시 3x3 컨볼루션 연산을 두 번 수행한 후 다시 2x2 최대풀링을 수행한다. 이후 4x4x16 영상을 Global Average Pooling 및 SoftMax 수행하여 최종

1x1x9 결과를 출력하는 구조로 구성되어 있다. 컨볼루션 연산은 모두 1x1 스트라이드를 적용하였으며 연산속도 향상을 위해 패딩을 적용하지 않았다. 연산량을 줄이기 위해 컨볼루션 레이어 필터 수를 최대한 줄인 모델로 오버피팅이 발생 가능한 환경이 아니기 때문에 드롭아웃을 적용하지 않았다.

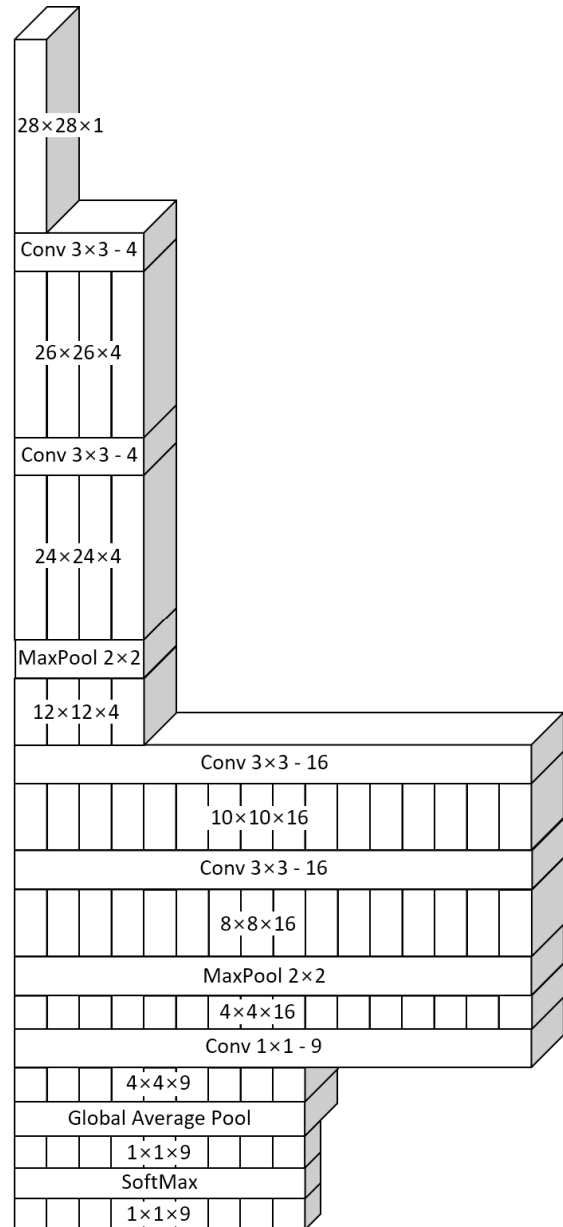


Fig. 5. Deep Convolution Neural Network Structure of Model D

이러한 모델 구조는 d'Acromont[11]의 기존 VGG 스타일을 따르는 방식에 비해 필터 수를 조정하여 연산량을 줄일 수 있으며, Kim[12]의 적외선 변형 최적화(IVO) 추가 연산 및 완전연결(FC)을 사용하지 않아 연산량을 줄일 수 있다.

이 모델 구조에서 VGG 스타일을 따르는 곳은 첫 번째 최대풀링 레이어 전 컨볼루션 연산 두 번 수행하는 곳과 첫 번째 최대풀링 레이어 후 컨볼루션 연산 두 번 수행하는 부분 밖에 없다. 여러 모델 변형 시 첫 번째 최대풀링 레이어를 기준으로 입력단과 출력단 컨볼루션 레이어의 필터 수를 각각 동일하게 적용하는 방식으로 변형하여 시험을 진행하였다.

Table 2. Model Configuration

A	B	C	D
Input 28×28 Gray Image			
Conv3×3-8	Conv3×3-6	Conv3×3-5	Conv3×3-4
Conv3×3-8	Conv3×3-6	Conv3×3-5	Conv3×3-4
MaxPool2×2			
Conv3×3-8	Conv3×3-12	Conv3×3-14	Conv3×3-16
Conv3×3-8	Conv3×3-12	Conv3×3-14	Conv3×3-16
MaxPool2×2			
Conv1×1-9			
Global Average Pool			
SoftMax			
Parameters 1,913	Parameters 2,475	Parameters 2,837	Parameters 3,253

Table 3. Number of Calculation of Models

A	B	C	D
Mul 476,064	Mul 372,600	Mul 337,932	Mul 314,640
Add 464,592	Add 362,976	Add 329,232	Add 306,864
MaxPool 1,280	MaxPool 1,056	MaxPool 944	MaxPool 832
Total 941,936	Total 736,632	Total 668,108	Total 622,336
Ratio 1.000	Ratio 0.782	Ratio 0.709	Ratio 0.661

표 2는 네 가지 모델들의 레이어 구성과 총 학습 파라미터 수를 나타내고 있다. 표 3은 네 가지 모델에서 Global Average Pooling 이전 레이어까지의 곱하기, 더하기, 최대풀링 연산량을 나타낸다.

모델 A는 임베디드 시스템에서 연산 가능한 수준으로 컨볼루션 레이어 필터 수를 처음으로 최적화한 모델로 연산량 최적화보다 학습 파라미터 수를 최소화하기 위해 조정된 모델이다. 최대풀링 레이어 기준으로 입력단과 출력단에 똑같이 8개 필터를 적용하였다. 모델 A는 학습 파라미터 수가 최소화되어 1,913개 파라미터와 연산량 941,936을 필요로 하였다. 모델 B는 최대풀링 레이어 기준으로 입력단 필터 6개, 출력단 필터 12개를 적용한 모델로 학습 파라미터 2,475개와 연산량 736,632를 필요로 했으며 모델 A에 비해 약 22% 연산량을 감소했다. 모델 C는 최대풀링 레이어 기준으로 입력단 필터 5개, 출력단 필터 14개를 적용한 모델로 학습 파라미터 2,837개와 연산량 668,108을 필요로 했으며 모델 A에 비해 약 29% 연산량

을 감소했다. 모델 D는 최대풀링 레이어 기준으로 입력단 필터 4개, 출력단 필터 16개를 적용한 모델로 학습 파라미터 3,253개와 연산량 622,336을 필요로 했으며 모델 A에 비해 약 34% 연산량을 감소했다. 최종적으로 모델 D를 선택하여 임베디드 시스템에 적용하였다.

3. Training

학습 환경은 다음과 같다. 딥러닝 라이브러리는 Tensorflow 2.4.1을 이용하였으며, 언어는 Python 3.8을 이용하였다.

컨볼루션 레이어는 다음과 같이 설정하였다. 파라미터 초기화 시 gloriot_uniform 옵션을 이용하였으며, 바이어스는 0으로 초기화하였고, 활성화함수(Activation Function)은 ReLU 함수를 적용하였다.

학습 옵션은 다음과 같이 설정하였다. 옵티마이저는 Adam을 이용하였으며, 손실함수(Loss Function)은 sparse_categorical_crossentropy를 이용하였고, 배치 크기는 32개, 확인용 데이터 비율 0.1, 학습 횟수 총 40회를 적용하였다.

학습 하드웨어 환경은 다음과 같다. CPU는 AMD Ryzen 9 5900X를 사용하였으며, 그래픽카드는 NVIDIA Geforce RTX 3090 24g를 사용하였으며, 시스템 메모리는 128g를 사용하였다. 윈도우10 Pro 64비트 시스템을 사용하였으며, CUDA 11.2.1 버전과 CuDNN 11.2 버전을 사용하였다.

그림 6은 모델 D의 학습과정으로 학습용 데이터 정확도 및 확인용 데이터 정확도를 나타내고 있다. 학습용 데이터 정확도와 확인용 데이터 정확도가 비슷한 경향으로 증가하여 96% 초반에 수렴하는 모습을 볼 수 있다. 학습용 데이터 손실함수 값과 확인용 데이터 손실함수 값이 비슷한 경향으로 약 0.12 값에 수렴되는 모습을 볼 수 있다. 학습용 데이터와 확인용 데이터에 기인한 정확도와 손실함수가 비슷한 경향으로 수렴하여 언더피팅이나 오버피팅 없이 학습이 수행되었음을 확인할 수 있다.

본 논문에서 제시하는 VGG 스타일 강화 모델은 입력 영상의 낮은 실제 해상도를 기반으로 배경과 표적 8종을 구분하는 모델로 연산량을 최대한 줄이기 위해 컨볼루션 레이어에 할당되는 필터 수를 언더피팅이 발생하지 않는 한 최대한 줄인 모델이다. 이런 이유로 모델 D의 경우 실제 학습 파라미터 수는 3,253개에 불과하다. 정확도를 유지하면서 한계까지 줄인 컨볼루션 레이어 필터 수와 연산량을 줄이기 위해 적용한 Global Average Pooling의 지역성으로 인해 파라미터 초기화 영향을 많이 받는 현상을

보였다. 학습 시 약 10~20번 시도에 한번 정도 약 95%가 넘는 정확도를 획득할 수 있었으며, 대부분 학습 시도에서 약 85% 정확도 획득에 그쳤다.

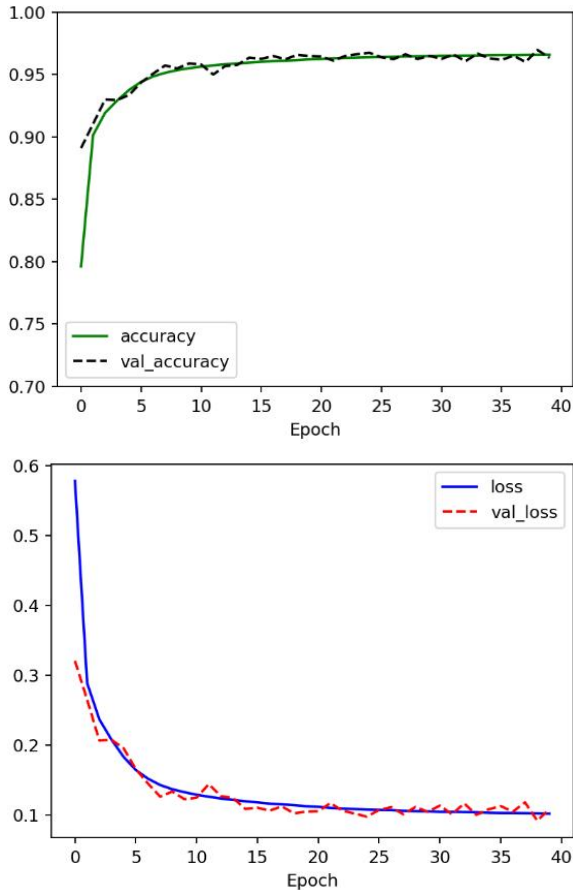


Fig. 6. Training Result of Model D (Accuracy, Validation Accuracy, Loss, Validation Loss)

4. Training Result

표 4와 같이 모델 A는 학습 정확도 0.937와 테스트 정확도 0.939, 모델 B는 학습 정확도 0.953과 테스트 정확도 0.957, 모델 C는 학습 정확도 0.956과 테스트 정확도 0.956, 모델 D는 학습 정확도 0.961과 테스트 정확도 0.963을 보였다.

모델 A에 비해 모델 D는 입력 영상을 기준으로 첫 번째 최대풀링 레이어 기준으로 입력단 필터 수가 8개에서 4개로 감소하였으며, 출력단 필터 수가 8개에서 16개로 증가되었다. 이는 기존 최대풀링 레이어 기준으로 필터 수를 2배씩 늘리는 VGG 스타일을 강화한 형태이다. 이러한 방식으로 최대풀링 입력단 필터 수를 줄이면 최대풀링 레이어 입력 영상 크기 28×28 및 26×26 기준으로 연산량이 절반으로 줄어들게 되며, 출력단의 필터수를 늘릴 경우 영상 크기 12×12 및 10×10 기준으로 연산량이 두배 늘어나게

된다. 이 경우 연산을 해야 하는 기준 영상 크기로 인해 최대풀링 입력단이 출력단보다 데이터량이 4배 크기 때문에 출력단 필터 수를 2배 늘리더라도 전체 연산량 34% 줄이면서도 정확도를 높일 수 있었다.

Table 4. Accuracy of Models

A	B	C	D
Training Acc 0.937	Training Acc 0.953	Training Acc 0.959	Training Acc 0.961
Test Acc 0.939	Test Acc 0.957	Test Acc 0.956	Test Acc 0.963

VGG 스타일을 강화한 형태로 전체 연산량을 줄이면서도 높은 정확도를 보였다. 이를 통해 다음과 같이 유추해볼 수 있다. 그림 2와 같은 저해상도 영상에서는 최대풀링 입력단 저레벨에서 실제로 필요로 하는 필터 종류가 적음을 알 수 있으며, 최대풀링 출력단 고레벨에서는 많은 필터 수를 실질적으로 필요로 하며, 이런 결과는 저해상도 영상에서는 저레벨 데이터가 크게 중요하지 않으며 추상화된 고레벨 추상화된 데이터 중요도가 높음을 알 수 있다.

5. C Code Porting

성능이 낮은 임베디드 시스템에 추론 시 발생하는 연산 속도 비교를 위해 모델의 추론 부분을 C 코드로 포팅하였으며, 표 2에서 제시된 컨볼루션 레이어, 최대풀링 레이어, Global Average Pooling, SoftMax 추론 연산 부분을 C 코드로 작성하였다.

실행 환경은 다음과 같다. CPU는 “AMD Ryzen 9 5900X”, 시스템 메모리는 128g를 사용하였다. 윈도우10 Pro 64비트 시스템을 사용하였으며, Visual Studio 2019 버전에 x64 모드, Release 모드를 사용하였다.

Table 5. Inference Lead Time of Models on ported C code

A	B	C	D
243 us	195 us	180 us	166 us

추론 연산 10,000회 수행 후 총 소요 시간을 연산횟수로 나누어 1회 추론 수행 시간을 계산하였다. 추론 시 포팅된 C 코드에서 모델 A는 243us, 모델 B는 195us, 모델 C는 180us, 모델 D는 166us 시간을 소요되었다. 모델 D가 모델 A에 비해 약 32% 시간을 덜 소비했으며 이는 컨볼루션 레이어 필터 수를 VGG 스타일을 강화하는 방향으로 조정하여 줄어든 연산량 34%가 반영된 결과이다.

IV. Conclusions

본 논문에서는 제안한 강화된 VGG 스타일 딥러닝 컨볼루션 신경망 모델을 이용하여 성능이 낮은 임베디드 시스템에 추론 가능한 분류 모델을 제안하였다. 최대풀링 레이어를 기준으로 8개와 8개 필터 수를 적용하여 파라미터 수를 최소화한 모델에 비해 4개와 16개 필터 수를 적용한 모델이 연산량을 34%를 감소하면서도 정확도를 93.7%에서 96.1%로 약 2.4%를 높이는 결과를 획득할 수 있었다. 이외에도 C 코드 포팅으로 줄어든 연산량 만큼 연산 수행시간도 약 32% 줄어든 것을 확인할 수 있었다. 이를 통해 파라미터 수 최소화를 통한 연산량 최적화가 이루어진 상태에서도 간단한 필터 수 비율 조절을 통해 연산성능 요구량을 줄이면서 최종 정확도를 높이는 것이 가능함을 확인할 수 있었다.

논문에 제시된 방법 이외에도 연산량을 줄이기 위해 MobileNet[8]의 Separable 컨볼루션 적용을 시도해 보았지만 필터 수를 32개 또는 64개 다수 적용하여도 최종 정확도 85% 이상 획득하기 어려웠다. 이는 MobileNet이 고 해상도 영상에 적합하도록 최적화되었기 때문에 나타나는 현상으로 추정된다.

본 논문에서 사용된 저해상도 적외선 영상은 차종을 겨우 분간할 수 있을 정도로 낮은 해상도 영상으로 최종 분류 정확도를 높이는 것은 쉬운 일이 아니다. 향후에는 연산량을 유지 혹은 소폭 증가하면서도 정확도를 크게 높일 수 있는 방법을 연구할 예정이다.

REFERENCES

- [1] D. A. Scribner, M. Kruer, and J. Killiany, "Infrared focal plane array technology," *Proc. IEEE*, vol. 79, no. 1, pp. 66-85, Jan 1991.
- [2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, Vol. 60, No. 6, pp 84-90, May. 2017.
- [3] Matthew D. Zeiler, Rob Fergus, "Visualizing and Understanding Convolutional Networks," *European Conference on Computer Vision*, pp. 818-833, Nov. 2014.
- [4] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition" *International Conference on Learning Representations*, Apr. 2015.
- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, "Going Deeper with Convolutions," *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2015.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition" *IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, Vol.1, 2016.
- [7] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, "Densely Connected Convolutional Networks," *Computer Vision and Pattern Recognition, Machine Learning*, Jan. 2018.
- [8] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *Computer Vision and Pattern Recognition*, Apr. 2017.
- [9] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *The IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018.
- [10] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, Jian Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile," *Computer Vision and Pattern Recognition*, Dec. 2017.
- [11] Antoine d'Acromont, Ronan Fablet, Alexandre Baussard, Guillaume Quin, "CNN-Based Target Recognition and Identification for Infrared Imaging in Defense Systems," *Multidisciplinary Digital Publishing Institute*, Vol.19, No.9, Apr. 2019.
- [12] Sungho Kim, Woo-Jin Song, So-Hyun Kim, "Infrared Variation Optimized Deep Convolutional Neural Network for Robust Automatic Ground Target Recognition," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Jul. 2017.
- [13] Jisoo Park, Jingdao Chen, Yong K. Cho, Dae Y. Kang, Byung J. Son, "CNN-Based Person Detection Using Infrared Images for Night-Time Intrusion Warning Systems," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Dec. 2019.
- [14] Aparna Akula, Anuj K Shah, Ripul Ghosh, "Deep Learning Approach for Human Action Recognition in Infrared Images," *Cognitive Systems Research*, Vol. 80, pp. 146-154, Aug. 2018.
- [15] Pilwon Kwak, Kiduck Kim, Hyochoong Bang, "Deep Transfer Learning Between Heterogeneous Data for Automatic Target Recognition," *Journal of Institute of Control, Robotics and Systems*, Vol. 24, No. 10, pp. 954-961, Oct. 2018.
- [16] Junhwan Ryu, Sungho Kim, "Data Driven Proposal and Deep Learning-based Small Infrared Drone Detection," *Institute of Control, Robotics and Systems*, Vol. 24, No. 12, pp. 1146-1151, Dec. 2018.
- [17] SungMok Yang, JinKyu Choe, "A Study on Improvement of Human Sensing System Reliability Using Thermal Image," *Korean Institute of Information Technology*, Vol. 17, No. 1, pp. 2015.

35-40, Jan. 2020.

- [18] Wonsik Oh, Ugwiyeon Lee, Jeongseok Oh, "Deep Learning(CNN) based Worker Detection on Infrared Radiation Image Analysis," Journal of the Korean Institute of Gas, Vol. 22, No. 6, pp. 8-15, 2018.

Authors



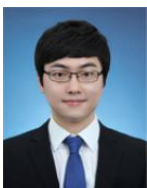
Yong-hee Hong received the B.S., M.S. degrees in Computer Science and Engineering from Soongsil University, Seoul, Korea, in 2009 and 2011, respectively. Mr. Hong joined LIG Nex1 company in 2011.

He is currently working on infrared camera. He is interested in image processing, developing infrared camera.



Sang-hun Jin received the B.S., and M.S. degrees in Electronics from Kyungpook National University, Korea, in 1993, 1995 respectively. Mr. Jin joined LIG Nex1 company in 2000. He is currently working

on search and tracking system. He is interested in image processing, developing search and tracking system based on infrared and visual camera.



Dae-hyeon Kim received the B.S., M.S. degrees in Electrical Engineering and Computer from Kyungpook National University, Daegu, Korea, in 2008 and 2010, respectively. Mr. Kim currently research

engineer at LIG Nex1 in Korea. His research interests include image signal processing, embedded system and infrared sensor tracking systems.



Ho-Jin Jhee received the B.S. degree in Electronics Engineering from Dongguk University, Seoul, Korea, in 1997 and the Ph.D degree in Electrical Engineering from University of Florida, Gainesville, Florida,

U.S.A in 2010. He is currently research engineer at LIG Nex1 in Korea. His research interests include statistical signal processing, image processing and applications of machine vision domain.