

# 딥 러닝을 이용한 흡연 검출 지능형 CCTV

## Deep Learning for Smoking Detection Intelligent CCTV

조혁준

### 요 약

라즈베리 파이와 같은 한정된 컴퓨팅 자원의 환경에서 카메라 모듈을 통해 입력되는 영상에서 흡연을 하는 경우 직접 학습한 딥 러닝 모델을 통해 식별할 수 있는 지능형 CCTV를 구현하고자 한다.

### 1. 서론

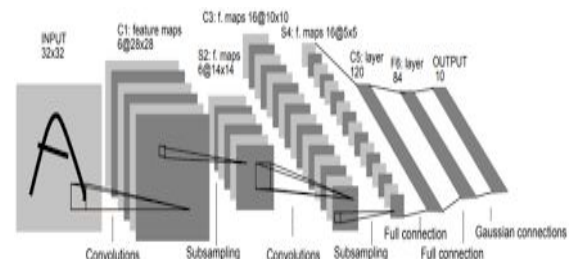
최근 금연 구역의 증가하는 추세를 볼 수 있다. 하지만 금연 구역에서 흡연을 하는 경우를 사람이 일일일 모두 찾아 낼 수 없으며 이러한 감시 비용에 큰 사회적 비용을 지출할 수 있다. 현재 Deep Learning 을 이용한 Object Detection 의 성능은 CNN(Convolution Neural Network) 을 통해 상당한 수준의 정확도를 보여준다. 이를 이용하여 라즈베리 파이와 같은 저렴한 비용으로 구현할 수 있는 시스템에서 담배 이미지들을 학습한 딥 러닝 모델을 통해 카메라에 입력받은 영상에서 흡연을할때 사용하는 담배를 검출할 수 있는 지능형 CCTV를 본 프로젝트에서 구현하기로 기획했다.

### 2. 기존연구

#### 2.1 CNN(Convolution Neural Network)

컨볼루션 신경망(CNN)은 심층 감독학습을 기반으로 하는 machine learning 모델이며, 적용력이 뛰어나고 국부적 특징 추출 및 분류에 강하다. 가중치 공유 구조 특징 때문에 컨볼루션 신경망 모델은 생물학적 신경망과 더욱 유사하게 설계되어으며 패턴인식 영역에서 탁월한 성과를 얻고있다. 컨볼루션 신경망은 convolution layer, pooling layer(or subsampling layer), fully connection layer 등으로 구성된다. 컨볼루션 신경망 모델에서 일반적으로

입력 층 input과 출력 층 output은 각각 1개로 고정되어 있고 convolution layer와 pooling layer를 여러 개로 구성할 수 있다. 전역 연결 층인 fully connection layer는 최종 출력층을 구성하기 위한 구조임을 볼 수 있다. Fully connected layer에서는 일반적으로 back propagation 알고리즘이나 back propagation 알고리즘의 단점을 보완한 gradient descent method나 wake-sleep 알고리즘을 적용한다. 컨볼루션 신경망은 특징을 보면 sparse weight를 통해 모델의 복잡도를 줄여주는 장점이 있고 parameter sharing을 통해 특정 가중치 그룹들은 가중치 값이 항상 같도록 변수를 공유하게 한다. 마지막으로 상위 sparse weight를 특정한 형태로 배치하였을 때, 주어진 입력 값의 변화에 대해 출력이 효율적으로 변화하는 방식이 동등하게 변화도록 한다. CNN의 기본 구조는 Figure 1과 같다.[1]



(Figure 1. CNN Structure)

## 2.2 Lightweight Deep Learning

알고리즘 경량화는 경량 딥러닝 알고리즘과 달리, 모델이 표현하는 다양한 파라미터의 크기를 줄이는 데 주목적을 가지고 있다. 파라미터가 가지는 표현력을 가능한 한 유지하면서 불필요한 가중치를 최대한 없애기 위한 방법들이다. 일반적인 딥러닝 모델은 과파라미터(Over-parameterization)화 되어 있기 때문에 모델이 가지는 가중치의 실제값이 아주 작을 경우, 모델의 정확도에 큰 영향을 미치지 못하므로(이를 모델이 작은 가중치에 대한 내성을 가진다고 표현함), 이 값을 모두 0으로 설정하여 마치가지치기(Pruning)를 수행하는 것과 같은 효과를 내는 가중치 가지치기(Weight Pruning)가 대표적이다. 다음으로, 일반적인 모델의 가중치는 부동소수점값을 가지지만, 이를 특정 비트 수로 줄이는 양자화(Quantization)를 통해 기존 딥러닝의 표현력을 유지하면서 실제 모델의 저장 크기는 줄이는 방법이 있다. 마지막으로, 0과 1로 표현하여 표현력을 많이 줄이지만, 정확도의 손실은 어느 정도 유지하면서 모델 저장 크기를 확연히 줄이는 이진화(Binarization) 기법 등이 있다. [2]

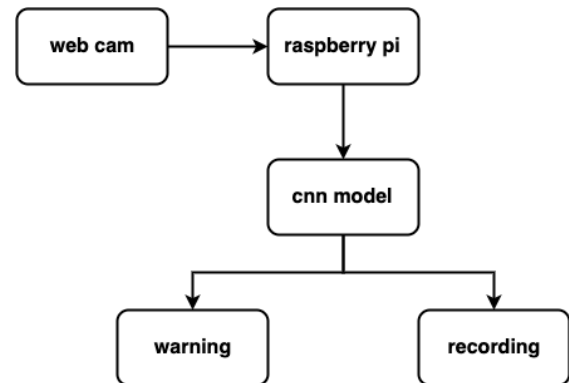
	접근방법	연구 방향
경량 알고리즘 연구	모델 구조 변경	전이 블록, 병목 구조, 밀집 블록 등 다양한 신규 계층 구조를 이용하여 파라미터 축소 및 모델 성능을 개선하는 연구(ResNet, DenseNet, SqueezeNet)
	합성곱 필터 변경	합성곱 신경망의 가장 큰 계산량을 요구하는 합성곱 필터의 연산을 효율적으로 줄이는 연구(MoblieNet, ShuffleNet)
	자동 모델 탐색	특정 요소(지연시간, 에너지 소모 등)가 주어진 경우, 경화 학습을 통해 최적 모델을 자동 탐색하는 연구(NetAdapt, MNasNet)
알고리즘 경량화 연구	모델 압축	가중치 가지치기, 양자화/이진화, 가중치 공유 기법을 통해 파라미터의 불필요한 표현력을 줄이는 연구(Deep Compression, XNOR-Net)
	지식 증류	학습된 기본 모델을 통해 새로운 모델의 생성 시 파라미터값을 활용하여 학습시간을 줄이는 연구(Knowledge Distillation, Transfer Learning)
	하드웨어 가속화	모바일 기기를 중심으로 뉴럴 프로세싱 유닛(NPU)을 통해 추론 속도를 향상시키는 연구
	모델 압축 자동 탐색	알고리즘 경량화 연구 중 일반적인 모델 압축 기법을 적용한 경화 학습 기법의 최적 모델 자동 탐색 연구(PocketFlow, AMC)

(Figure 2. 경량 딥러닝(Lightweight Deep Learning) 연구 동향)

## 3. 시스템 모델

### 3.1 시스템 구조 및 설계

먼저 웹캠과 같은 카메라 모듈을 통해 입력되는 영상을 직접 학습시킨 딥러닝 모델을 통해 추론하여 흡연 상황을 검출한다. 검출이 되었을 경우 경고 기능과 녹화 기능을 사용할 수 있는 시스템을 Figure 3 에서 볼 수 있는 것과 같이 설계하였다.



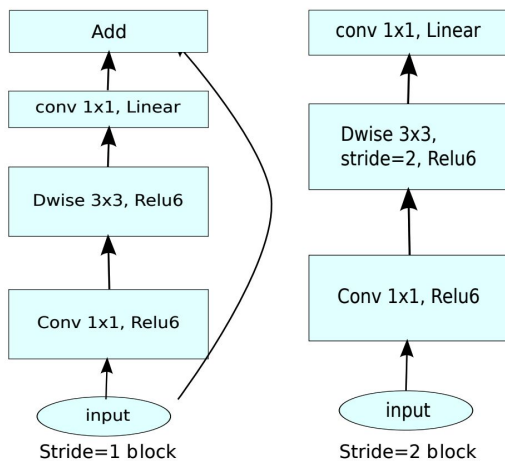
(Figure 3. System Diagram)

기존 연구들의 성능저하 이슈는 GPU의 메모리, 성능의 문제로부터 발생한다는 점에서 착안하여 추론단계의 퍼포먼스를 최대화 하기 위하여 모델의 메모리 사용량을 최소화 한다. 이를 위하여 Tensorflow로 학습 된 모델을 8-bit 양자화된 TFLite 형식으로 변환하여 사용한다. 또한 라즈베리 파이의 성능을 고려하여 모델에서 사용되는 데이터의 형식을 가공하여 Latency Delay를 낮춘다. 때에 따라서 메모리의 사용량을 최소화 하면 정확도가 떨어질 수 있는데 해당 부분을 잘 조율하기위해 다양한 학습 모델을 연구하고 선정하였다.

## 4. 프로젝트 내용

본 프로젝트는 Tensorflow를 이용하여 생성된 모델을 TF-Lite 형식으로 변환하여, 변환된 파일들을 라즈베리 파이 환경에 이식하는것을 목표로 한다. 먼저 데이터셋을 구성한뒤 labeling 과정을

거친다. labeling 과정에서 발생한 각각의 xml형식의 파일을 csv의 포맷으로 통합한다. 이미지와 labeling 데이터를 tensorflow을 통해 학습하기 위해 tfrecord 형식의 데이터셋으로 구성한다. 이때, train 용도와 validation 을 구분하여 구성한다. 다음 모델은 TF-Slim을 사용하여 학습하며, 이 때 이용하는 모델은 Single Shot Detector를 적용한 MobileNetV2 모델이다.

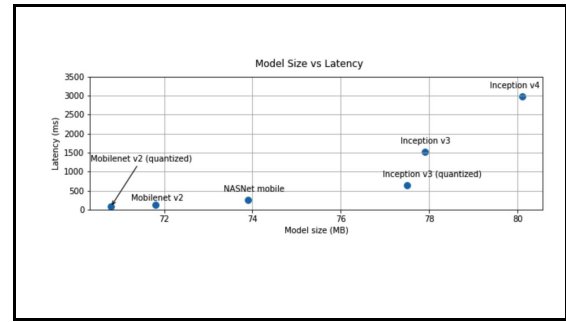


(Figure 4. MobileNetV2)

Size	MobileNetV1	MobileNetV2	ShuffleNet (2x,g=3)
112x112	64/1600	16/400	32/800
56x56	128/800	32/200	48/300
28x28	256/400	64/100	400/600K
14x14	512/200	160/62	800/310
7x7	1024/199	320/32	1600/156
1x1	1024/2	1280/2	1600/3
max	1600K	400K	600K

(Figure 5. Memory Usage Table)

해당 모델은 이전 버전의 V1, ShuffleNet 보다 채널수, 메모리 부분에서 가장 적은 수를 차지하기 때문에 제한적인 컴퓨팅 리소스를 가진 라즈베리 파이에서 적합하다고 판단하였고 또한 실시간으로 입력되는 디지털 영상 데이터를 처리하기위해 Figure 6.에서 Latency가 낮은 모델을 사용하는것이 합리적이라 생각하였다.



(Figure 6. Latency vs Model Size)

학습 된 모델은 proto buffer 형태로 생성되는데 모델의 사이즈를 줄이 위해 tflite 확장자로 변경 한 후 라즈베리 파이에 이식 하였다. 설계된 데이터셋은 Figure 7. 에서 확인할 수 있다.

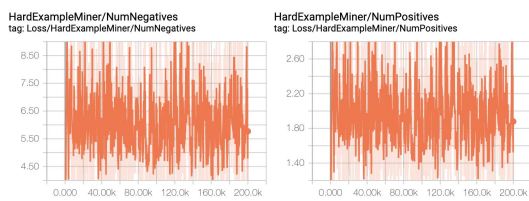
Class	Trained Image	Validation Image
smoking	780	70

(Figure 7. Datasets)

다양한 상황에서 담배를 검출하기 위해 먼저 구글에서 길거리에서 흡연하는 사람이 피고있는 담배 사진과 직접 흡연하며 촬영한 담배사진으로 직접 labeling 과정을 거쳐 구성하였다. 특히 다양한 앵글과 담배의 상태를 고려하여 데이터셋을 선정하였으며 전체 데이터셋 집합에서 10% 가량의 데이터셋을 검증집합으로 분류하였다. 이러한 데이터셋을 가지고 GPU를 이용하여 학습을 수행하였다.

## 5. 결과

먼저 학습과정에서 발생한 loss value 의 값들은 다음과 같다. Figure 8.에서 불규칙한 그래프를 볼 수 있는데 데이터셋의 일관된 데이터 상황을 가지지 못하여 발생한 결과로 추정된다.



(Figure 8. Loss Value Graph)

직접 데이터셋을 모아 라벨링을 거친 후 생성한 모델을 기반으로 물체를 인식하는 실험을 하기위해 직접 흡연하는 과정을 촬영하여 테스트하였다. Figure 9. 는 다양한 각도에서 촬영한 영상을 확인할 수 있다.



(Figure 9. Detection Test)

테스트를 수행한 결과 다양한 각도에서 담배를 검출 할 수 있었다. 하지만 대부분의 영상에서는 적절한 위치를 찾지 못하였으며 이러한 부분은 전체 이미지에서 차지하는 담배 이미지의 크기가 차지하는 부분이 5% 미만인것을 확인 할 수 있었다.

## 6. 결론 및 향후 연구

데이터셋을 직접 구성하여 학습한 딥러닝 모델을 라즈베리 파이를 환경에서 적용하여 지능형 CCTV를 설계하고 구현하였다. 이러한 시스템을 통해 금연 구역을 촬영하는 CCTV를 설치할시 사용자는 일일이 CCTV를 들여다 볼

필요가 없어지며 본 시스템에서 제공되는 경보와 녹화 기능을 이용하여 금연 구역에서 흡연을 하는 행위를 방지할 것으로 기대된다. 하지만 저조한 성능을 보완하기 위해 모델을 설계하는 과정에서 데이터셋을 다양한 상황과 환경에서 구성하고 실제 흡연 상황에서 담배는 전체 이미지에서 5% 미만의 크기를 가지고있어서 모델을 통해 추론하는데 어려움이있다. 이를 보완하기위해 먼저 얼굴과 손 부분을 cropping 하는 과정을 거쳐 resizing 된 이미지를 다각화하여 담배를 검출하는 모델의 입력 파라미터로 적용할 시 더 높은 추론이 가능하게 될 것으로 판단된다.

## 참고 문헌

- [1] Ga-Ae Ryu, Kwan-Hee Yoo, "Application of Manufacturing Process Data Classification Using Image Data based CNN" Journal of Information Technology and Architecture Vol. 15. No. 3, September 2018, Pages 337-343
- [2] 이용주, 문용혁, 박준용, 민욱기 "Recent R&D Trends for Lightweight Deep Learning":2019 Electronics and Telecommunications Trends
- [3] 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":arXiv:1704.04861v1 [cs.CV] 17 Apr 2017
- [4] Joseph Redmon, Anelia Angelova, "Real-Time Grasp Detection Using Convolutional Neural Networks"