

# 이미지 분석과 딥 러닝을 통한 영유아 위험물 탐지<sup>1</sup>

김휘준, 박길섭, 서영학, 김경섭  
 충남대학교 컴퓨터공학과

e-mail: joon4141@gmail.com, eric6679@naver.com, ox4443@naver.com, sclkim@cnu.ac.kr

## Detection of Dangerous Things to Infants through Image Analysis and Deep Learning

Kyung-Sup Kim, Hui-Joon Kim, Kil-Seop Park, Yeong-Hak Seo  
 Dept of Computer Engineering, Chungnam National University

### 요 약

본 논문은 이미지 탐지 모델인 Faster R-CNN 을 통해 영유아가 존재하는 어린이 집, 공원, 놀이터, 거실 등의 2D 이미지를 읽어 영유아에게 위험이 되는 요소를 인식해 위험상황을 감지하는 시스템을 구현하였다. 실생활에서 쉽게 구할 수 있는 데이터를 바탕으로 탐지 모델을 구현 했으며 현재 머신 러닝 분야가 음성인식과 행위데이터를 기반으로 상용화 되어있는 반면 본 모델은 이미지를 데이터로 한 탐지 모델이 다양한 서비스 분야에서 활용 될 수 있음을 보여준다.

### Abstract

In this paper, we implemented a system to detect dangerous situations by recognizing the dangerous elements for infants by reading 2D images of children's houses, parks, playgrounds, and living rooms where infants are present through Faster R-CNN. We have implemented a detection model based on data that can be easily obtained from real life. Currently, machine learning is commercialized based on speech recognition and behavior data. However, this model can be applied to various service fields Respectively.

### 1. 서론

2000 년대에 들어서면서 기계 학습의 진보는 연구 분야 및 상업적 응용 분야의 기반으로 인기를 얻어왔다. 기계학습은 우리가 사물을 인식하고 분석하는 방법을 컴퓨터 스스로 학습하는 데 도움을 주었으며 다양한 산업 및 상업적 응용 분야에서 이가 필요하다. 특히 AWS 사의 ECHO, SKT 사의 NUGU 등의 전자비서는 음성인식을 바탕으로 실생활에 편의를 제공하며 최근 4 차 산업의 자율주행 자동차 분야에서 딥러닝 기술 적용에 대한 연구가 활발하다.

현재 서비스업 분야에서 사용되고 있는 딥러닝 기술은 음성인식과 자연어 처리기술이 대부분을 이루고 있다. 본 논문에서는 딥러닝 기술을 이용한 이미지 인식 모델을 활용하여 상황 탐지 모델을 구현하였다. 특히, 영유아를 대상으로 주변에 널린 물체가 위험물이 될 수 있는지에 대한 상황을 설정했다.

한국 소비자자원(CISS)에서 제출한 보고서를[1] 통해 수집된 1~3 세 영유아 안전사고 분석 보고서에 따르면 안전사고의 80.4%가 주택에서 일어난다. 특히 낙하, 충돌, 삼킬 수 있는 물체에서 사고가 자주 발생하며

[표 1] 시나리오를 예방하기 위해 해당 위험 요소를 학습시키고 이를 토대로 이미지를 통해 영유아에게 위험한 상황을 감지하는 시스템을 만드는 것을 목적으로 한다.

걸음마기 위해유형별 현황

단위: 건, %

순위	위해유형	주요 원인품목(소분류)	건수	비율
1	미끄러짐/넘어짐	바닥재, 계단 등	10,301	27.5
2	부딪힘	거실장 및 TV장, 침대, 서랍장 등	9,948	26.6
3	추락	침대, 소파, 식탁 등	6,064	16.2
4	놀람/끼임	문, 기타 승용물, 엘리베이터 등	2,978	8.0
5	이물질 삼킴/흡인	구슬, 완구류, 조립용 블록, 스티커, 공류 등	2,892	7.7
6	베임/찔림	면봉, 눈썹용 칼, 가정용 칼, 가위 등	1,763	4.7
7	고온물체에 의한 위해	정수기, 전기밥솥, 다리미 등	1,419	3.8
8	식품 섭취로 인한 위해	우유, 달걀, 캔디 등	668	1.8
9	충돌/추돌 등 물리적 충격	거울, 서랍장, 행거 등	430	1.1
10	동물에 의한 상해	애완견	242	0.6
기타			714	2.0
합계			37,419	100.0

[표 1] 17년 상반기 영유아 위험사고유형

본 시스템은 단일 2D 이미지를 탐지하므로 카메라가 아닌 일반적인 캠, CCTV 등이 설치되어있는 공공

<sup>1</sup> 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW 중심대학지원사업의 연구결과로 수행되었음 (2015-0-00930)

시설이나, 어린이집, 공원, 놀이터, 거실 같은 일반적인 환경에서도 활용할 수 있다.

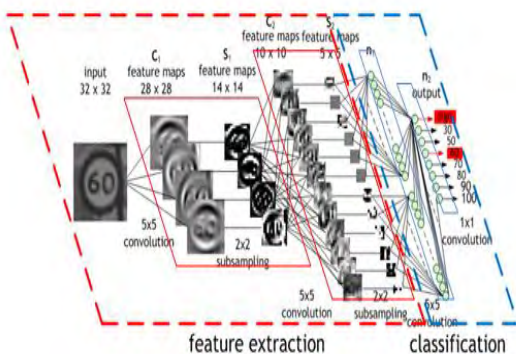
어린이들은 특성상 잠시만 한 눈을 팔아도 위험 상황에 처할 수 있기 때문에 이를 예방하기 위해 다양한 센서를 기반으로 한 시스템과 영상처리 알고리즘을 이용한 연구 사례가 다양하다 하지만 센서 기반의 모니터링은 몸에 늘 부착해야 하므로 사용성이 떨어진다 단점이 있으며 영상처리 알고리즘의 경우 각 물체별로 알고리즘을 다르게 적용하기 때문에 다양한 물체의 인식에 활용되기가 어렵다. 따라서 본 논문에서는 다양한 물체의 인식이 가능한 딥러닝 알고리즘을 이용하여 구현하였다.

본 논문은 영유아 위험 탐지 모델을 구성하는 훈련 모듈인 Faster R-CNN[2]에 대한 설명과 신경망을 구성하는 CNN에 대한 설명을 서술했으며 시스템 설계에 대한 내용이 서술되어 있다. 시스템 설계는 데이터 수집 및 전처리, 훈련, 시각화 순으로 이루어져 있으며 마지막으로 모델 평가에 관하여 서술되어 있다.

## 2. 연구 배경

### 2.1 CNN(Convolutional Neural Networks)[4]

이미지를 훈련하고 분류하여 고수준의 추상화된 정보를 추출하거나 그림을 그리는 등 다양한 영상처리, 컴퓨터 비전 분야에서 많이 연구되고 있는 인공지능망의 한 종류이다. CNN은 일반적으로 3 종류의 계층 Convolution 계층, Pooling 계층, Feed forward 계층이 있다. Convolution 계층은 합성 곱 연산을 이용하여 의미 있는 특징들을 추출하기 위한 층이다. Pooling 계층은 이미지에 존재하는 많은 픽셀의 특징을 줄이기 위해 Sub Sampling을 하는 층이다. Sub Sampling은 화면의 크기를 줄이는 과정인데 Sub Sampling을 하는 이유는 차원이 증가하면 그것을 표현하기 위한 데이터가 기하급수적으로 증가하기 때문에 데이터의 의미를 제대로 표출할 수 없다. 그러므로 특징을 추려내기 위해서 서브 Sub Sampling 과정을 진행한다. Feed forward 계층은 Convolution 계층과 Pooling 계층에서 나온 특징들을 추출하여서 분류를 해주는 층이다. CNN은 위와 같은 계층을 이용해 특징들을 추출하여 물체를 인식한다.



[그림 1] CNN의 일반적인 구조

### 2.2 Faster R-CNN(Faster Region-based Convolutional Neural Networks)[2]

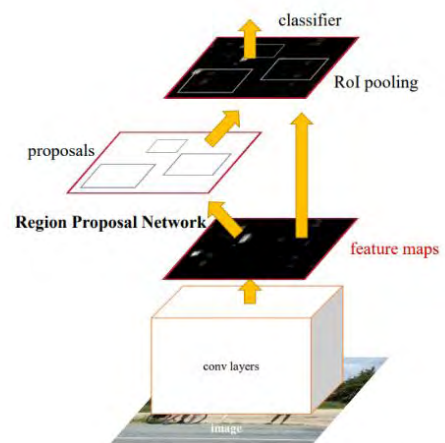
이미지 안에 어떤 물체들이 들어있는지 구분하기 위해 이미지 분류기 앞에 이미지 영역을 찾아내는 탐지 영역을 합친 모델이다. 탐지 영역은 특징 맵 추출과 RPN(Region Proposal Network), RoI Pooling 단계로 나누어져 있다.

특징 맵 추출 단계는 여러 사물이 포함된 복합 이미지에서 Convolution 계층을 이용해 특징 맵을 추출한다. 이후 생성된 특징 맵을 RPN에 통과시켜 경계 상자 점수와 좌표를 추출한다. 경계 상자 점수와 좌표는 특징 맵에서 슬라이딩 윈도우를 이용해 분류한다. 이때 1x1 필터와 사전에 정의된 k개의 anchor 상자를 사용하여 경계 상자와 경계 상자가 물체인지를 학습한다.

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

(1)은 RPN의 손실 함수로 물체 가능성을 학습하기 위한  $L_{cls}$ 와 경계 상자의 좌표를 학습하기 위한  $L_{reg}$ 로 이루어져 있다. 여기서  $p_i$ 는 예측된 물체 확률값을 의미하며  $p_i^*$  경우 anchor가 물체일 경우 1, 물체가 아닐 경우 0의 값을 갖는다. 그 때문에 손실 함수로 two-class Softmax나 logistic regression을 사용한다.  $t_i$ 는 예측된 경계 상자의 값을 의미하며 x 좌표, y 좌표, 폭, 높이 값을 가진다.  $t_i^*$ 은 물체의 경계 상자의 값을 가지며 손실 함수로 smooth  $L_1$ 을 사용한다.

얻어진 상자 정보를 Fast R-CNN[3]의 RoI(Region of Interest) Pooling layer를 통해 구해진 각 상자들을 분류해준다. 얻어진 경계 상자는 크기가 일정하지 않기 때문에 max pooling 계층을 이용하여 일정한 사이즈로 만든 뒤 물체와 상자를 분류해준다.

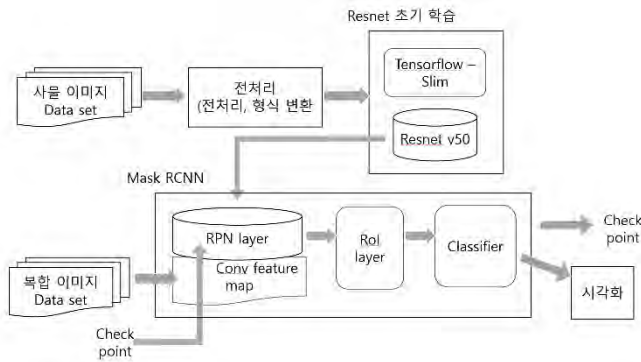


[그림 2] RPN 구조

Faster R-CNN의 훈련은 4 단계의 교대훈련으로 이루어져 있다. 1 단계는 pre-trained 된 데이터 셋을 통해 RPN을 훈련한다. 2 단계는 1 단계에서 제안된 박스를 가지고 RoI Layer를 훈련시킨다. 3 단계는 2 단계를 바탕으로 RPN을 훈련시키며 이때 feature map을 추출하는 layer는 고정된다. 마지막으로 3 단계 결과를 가지고 Roi Polling Layer를 훈련한다.

### 3. 시스템 설계

본 위험물 탐지 시스템을 위해 아이, 어른, 그리고 위험물들이 분포된 복합 이미지에 라벨링을 하여 데이터 셋을 만드는 전처리 과정을 거쳤다. 전 처리된 데이터 셋을 훈련시킨 뒤 Faster R-CNN(Faster Region-based Convolutional Neural Networks)[2] 모델을 사용하여 학습시켰으며 이미지 분류 모델로 resNet50[5]을 사용해 훈련 횟수를 감소시켰다. 훈련된 모델을 바탕으로 영유아가 있는 환경에서 영유아, 어른, 위험물 탐지하고 위험 상황을 보여주는 시각화 모델로 이루어져 있다.



[그림 3] 시스템 구성도

#### 3.1 학습 데이터

훈련 반복횟수를 감소시키기 위해 사전 학습을 위한 데이터를 수집하였다. [표 1]의 영유아 사고 유형에 따라 낙하 요소(소파, 의자, 테이블), 충돌 요소(칼), 영유아, 성인 이미지를 학습데이터로 정의했다. 웹상에서 각각 600 개의 2D 이미지를 사용하였으며 훈련 데이터 500 개, 검증 데이터 50 개, 시험 데이터 50 개로

데이터 셋을 구성하였다.

여러 사물이 들어있는 복합이미지는 영유아가 많은 어린이집의 배경으로 한 250 개의 2D 이미지 사용하였으며 훈련 데이터 200 개, 검증 데이터 20 개, 시험 데이터 30 개로 데이터 셋을 구성하였다.

#### 3.2 데이터 전처리

사전 학습을 위한 데이터 셋은 각 이미지 별로 라벨을 달아 Tensorflow의 데이터 셋 형식인 Tfrecord로 만들었다. 복합이미지의 경우 각 이미지별로 내부에 존재하는 사물의 이름과 개수 그리고 이미지 내부에서 픽셀 위치를 라벨링 하였으며 TFrecord 형식으로 만들었다.

#### 3.3 학습 모듈

##### 3.3.1 Tensorflow Slim

Tensorflow Slim은 이미지 분류에 사용되는 딥러닝 CNN 모델을 제공한다. 지원되는 모델은 AlexNet, VGG, ResNet[5], Inception-V3 등이 있으며 네트워크 망뿐만 아니라 새로운 데이터 셋에 대한 Fine Tune 기능도 제공한다. 본 시스템에서는 Tensorflow Slim에서 지원하는 ResNet\_V1\_50 모델을 사용하여 사전 학습을 진행하였으며 총 10000 번의 반복학습을 진행하였다.

##### 3.3.2 Faster R-CNN

Faster R-CNN은 Tensorflow로 구현이 되었으며 전 처리된 이미지와 반복학습을 줄이기 위한 사전훈련데이터를 입력으로 받아 복합이미지에서 사물을 인지한다. 복합이미지에서 사물을 인지하기 위한 딥러닝 알고리즘으로는 R-CNN, Fast R-CNN, Faster R-CNN 등이 있다. R-CNN과 Fast R-CNN의 경우 사물의 위치를 학습하는 데 있어 영상처리 알고리즘을 사용하기 때문에 학습시간이 길어지며 예측모델에서의 시간이 길어진다. 하지만 Faster R-CNN의 경우 전 모델의 영상처리 부분을 CNN으로 대체하여 학습시간이 크게 단축되며 예측시간도 짧기 때문에 선정하였다.

훈련이 시작되면 Faster R-CNN에 정의된 Layer들을 사전에 ResNet으로 학습시킨 데이터를 가지고 초기화를 한다. 이후 데이터 셋 중 훈련 데이터 셋 부분을 랜덤하게 큐에 넣으며 각각의 이미지에 대하여 훈련



(a) 원본 이미지



(b) 라벨 이미지



(c) 결과 이미지

[그림 4]

을 진행하며 훈련이 완료되었을 시 검증 데이터를 이용해 모델을 검증하고 정확도를 측정한다. 사전 데이터를 사용하는 것은 첫 회에 한정되면 이후 추가로 훈련하고자 할 경우 훈련된 데이터를 바탕으로 진행이 된다. 훈련 결과는 Ckpt 파일 형식으로 저장되며 10000 번의 학습마다 저장된다.

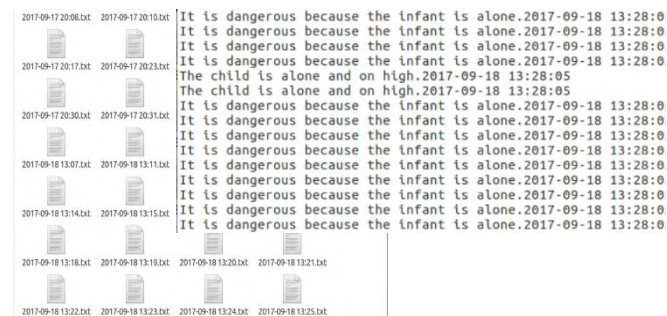
### 3.4 시각화

Faster R-CNN 신경망을 통해 훈련된 계층을 사용하여 얻은 물체 인식 정보(사물 이름, 상자 좌표, 정확도)를 바탕으로 [표 2]와 같이 위험 시나리오를 구성했다. 인식된 이미지의 종류, 성인의 유무를 기준으로 위험 정도를 정의했으며 낙하 사고의 아이와 영유아와 물체 간의 픽셀을 기준으로 위험 정도를 정의하였다. 결과는 물체 인식 사진이 저장되며 위험 상황의 경우 로그 기록이 남는다.

## 4. 시뮬레이션 결과

[그림 4]는 원본 이미지(a), 라벨링 한 이미지(b), 결과 이미지(c)를 이다. 결과 분석을 위해 Faster R-CNN 훈련에 쓰이지 않은 (a)를 위 시스템에 적용해 보았다. (b)와 (c)를 비교해보면 박스의 개수와 위치가 차이가 있음을 알 수 있다. 이는 시스템이 물체를 인식할 때 RoI layer 에서 잡아준 위치 정보가 겹칠 수 있기 때문이다. 물체 인식이 겹친다고 해도 다른 위험물 간의 관계로 위험물을 감지하는 데 영향을 주지 않기 때문에 문제가 되지 않는다.

우리는 범용적인 시설에서의 적용을 목표로 했기 때문에 이미지 분류 모델에 선 학습을 한 후 본 시스템에 적용하여 복합이미지를 학습하였다. 이 결과를 영유아와 성인이 같은 이미지에 존재하는지 아닌지를 바탕으로 위험도를 판단했을 때 50 개의 검증 데이터 셋에 적용했을 때 92%의 정확도를 보였다.



[그림 4] 로그 이미지

## 5. 결론

본 논문에서는 Faster R-CNN 모델과 이를 구성하는 resNet 신경망을 적용하여 물체의 위치를 탐지하였다. 먼저 훈련 초기값을 위해 복합이미지가 아닌 단일 이미지를 ResNet 신경망을 적용해 학습시켰다. 다음 단계로 복합 이미지 내부에 존재하는 물체의 종류와 위치를 라벨링 하여 학습 및 시험을 위한 데이터 셋을

구성하였다. Faster R-CNN 모델에 위험물과 영유아의 이미지를 학습시킨 resNet 모델을 적용한 후 영유아, 성인, 그리고 위험물이 라벨링 된 복합 이미지를 학습시켰다. 학습 후 다른 이미지 셋을 넣었을 때 정확하게 아이, 어른, 위험물의 위치를 파악하였다.

테스트 과정에서 위 모델이 복합 이미지를 인식한 결과와 사람이 물체를 확인한 결과는 비슷한 수준으로 나타나는 걸 확인할 수 있었다. 이 결과를 바탕으로 위험 상황인 경우 로그를 남겼으며 인식한 이미지에 박스를 추가하여 시각화 하였다.

앞으로 위험물 감지의 여러 방법을 실험할 예정이다. 특히 신경망의 분류기 뒤에 결과(물체 종류, 위치 정보)를 입력으로 하는 RNN 모델을 넣어 2D 이미지의 위험 상황 자체를 학습하는 신경망을 구상 중이다. 이를 이용하면 아이에게 위험한 거리 요소를 영상처리 방법을 이용하지 않고 학습할 수 있다.

현재 각 분야에 머신 러닝이 적용되기 어려운 이유 중 하나는 사용할 수 있는 데이터의 부족에 있다. 이를 극복하기 위하여 이미지 생성을 이용한 데이터 셋 구성에 관하여 연구할 예정이며 더 많은 데이터를 바탕으로 실용성을 검증해 실제 적용 가능한 기술을 개발할 예정이다.

## 참고문헌

- [1] 한국 소비자 안전센터 소비자 안전국 위해정보팀, “어린이 안전사고 동향 분석 보고서,” 한국소비자원, pp. 18-22, 2017.
- [2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks” arXiv:1506.01497v3 [cs.CV] 6 Jan 2016
- [3] R. Girshick. Fast R-CNN. In ICCV, 2015.
- [4] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton “ImageNet Classification with Deep Convolutional Neural Networks” 2012
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, Microsoft Research “Deep Residual Learning for Image Recognition” arXiv:1512.03385v1 [cs.CV] 10 Dec 2015