

R-FCN과 Transfer Learning 기법을 이용한 영상기반 건설 안전모 자동 탐지

박상윤* · 윤상현** · 허 준***

Park, Sangyoon*, Yoon, Sanghyun**, Heo, Joon***

Image-Based Automatic Detection of Construction Helmets Using R-FCN and Transfer Learning

ABSTRACT

In Korea, the construction industry has been known to have the highest risk of safety accidents compared to other industries. Therefore, in order to improve safety in the construction industry, several researches have been carried out from the past. This study aims at improving safety of labors in construction site by constructing an effective automatic safety helmet detection system using object detection algorithm based on image data of construction field. Deep learning was conducted using Region-based Fully Convolutional Network (R-FCN) which is one of the object detection algorithms based on Convolutional Neural Network (CNN) with Transfer Learning technique. Learning was conducted with 1089 images including human and safety helmet collected from ImageNet and the mean Average Precision (mAP) of the human and the safety helmet was measured as 0.86 and 0.83, respectively.

Key words : Construction safety, Object detection, Deep learning, Neural network

초 록

대한민국에서 건설업은 타 업종들과 비교하여 안전사고의 위험성이 가장 높게 나타난다. 따라서 건설업 내 안전성 향상을 도모하기 위해 여러 연구가 예전부터 진행이 되어 왔고, 본 연구에선 건설현장 영상 데이터를 기반으로 물체 탐지 및 분류 알고리즘을 이용해서 효과적인 안전모 자동 탐지 시스템을 구축하여 건설현장 노동자들의 안전성 향상에 기여하고자 한다. 본 연구에서 사용된 알고리즘은 Convolutional Neural Network (CNN) 기반의 물체 탐지 및 분류 알고리즘인 Region-based Fully Convolutional Networks (R-FCN)이고 이를 Transfer Learning 기법을 사용하여 딥러닝을 실시하였다. ImageNet에서 수집한 1089장의 사람과 안전모가 포함된 영상으로 학습을 시행하였고 그 결과, 사람과 안전모의 mean Average Precision (mAP)은 각각 0.86, 0.83로 측정되었다.

검색어 : 건설안전, 물체 탐지, 딥러닝, 인공신경망

* 정희원 · 연세대학교 건설환경공학과 석사과정 (Yonsei University · parksangyoon@naver.com)

** 연세대학교 건설환경공학과 박사과정 (Yonsei University · yoonssa@yonsei.com)

*** 종신회원 · 교신저자 · 연세대학교 건설환경공학과 정교수 (Corresponding Author · Yonsei University · jheo@yonsei.ac.kr)

Received January 8, 2019/ revised April 8, 2019/ accepted April 10, 2019

1. 서론

1.1 연구 배경

건설업에서 건설노동자들이 위험한 환경에 노출되어 있음에도 불구하고 필수적인 안전 수칙들조차 잘 지켜지지 않고 있다. 이로 인해 크고 작은 안전사고가 자주 발생하고 있으며 그 수는 매년 증가하고 있다(You et al., 2017). 국내 산업재해 발생 현황 보고서에 의하면 ‘보호구 착용 상태 및 성능 불량’은 전체 안전사고 원인의 13.3 %으로 2위를 차지했다. 상해 부위별 사고 사망자 현황을 보면 머리(48.4 %), 몸통(26.9 %), 다발성 신체 부위(13.4 %) 순으로 많았으며 상해 종류별로는 두개(頭蓋) 내 손상(외상성)이 45.8 %로 가장 많이 차지하였다. 안전모 등 개인 보호 장비를 착용하지 않은 경우의 안전사고 발생률은 88.9 %로, 개인 보호 장구를 착용했을 때의 안전사고 발생률인 11.1 %보다 8배나 높게 분석되었다(Occupational Safety and Health Research Institute, 2016). 종합하면, 현재 많은 건설현장에서 안전모를 착용하는 것만으로도 충분히 예방 가능한 재해가 많이 발생하고 있으나 건설노동자들의 안전불감증으로 인해 안전사고로 이어지는 상황이다. 그러나 현재 대부분의 건설현장에서 안전관리자가 건설노동자들의 안전모 착용 여부를 지속적으로 감시하는 것은 힘든 상황이다. 이러한 문제를 해결하기 위하여 현장에 설치되어있는 CCTV와 같은 카메라 장치에서 획득한 영상 데이터를 기반으로 딥러닝 알고리즘을 적용하여 건설노동자들의 안전모를 자동으로 탐지하는 방법을 제안하고자 한다.

1.2 선행 연구

안전모에 대한 탐지를 영상 인식으로 수행하는 방법은 이미 해외 컴퓨터 비전 분야에서 지속적인 연구가 되었으며, 기하학적인 특징을 기반으로 기계학습을 같이 적용하는 선행 연구가 진행되었다.

Wen et al.(2003)은 Hough 변환(Illingworth and Kittler, 1988)을 기반으로 해서 원과 호(Circle Arc) 탐지 방법을 제안했다. 하지만 이 방법은 기하학적으로만 특징만을 사용하므로 사람의 머리가 안전모로 잘못 인식될 확률이 매우 높다. 이러한 한계점을 해결하기 위해 기계학습 알고리즘을 적용하여 검출률을 높이는 여러 연구가 진행되었다. Silva et al.(2013)는 Histogram of Oriented Gradients (HOGs)와 Circular Hough Transform (Yuen et al., 1990) 사용해서 특징을 추출하고 추출된 특징을 기계학습 알고리즘인 Support Vector Machine (SVM)을 이용해서 분류했다. Rubaiyat et al.(2016)은 이러한 과정에 안전모의 색을 인식하는 단계를 추가하여 검출률을 높였다. 하지만 HOG와 SVM를 이용한 2가지 단계로 분류해서 진행하는 방법은 영상 한 장당, 연산 시간이 10분 이상 걸려 실시간 탐지에는 적합하지 않고(Memarzadeh

et al., 2013) 노동자들이 HOG의 특징 템플릿에 맞지 않는 행동을 취하면 안전모가 정상적으로 인식되지 않는 문제점을 보인다(Fang et al., 2018).

최근 물체 탐지 분야(Object Detection)에서는 기존의 알고리즘과 비교하여 압도적인 성능 차이를 보이는 신경망(Neural Network, NN)을 기반으로 물체를 탐지 및 분류를 하는 방법이 대두되고 있다(Vishnu et al., 2017). 그 중에서도 회선신경망(Convolutional Neural Network, CNN)을 이용한 물체탐지 알고리즘은 Region-based Convolutional Neural Network (R-CNN)을 시작으로(Girshick et al., 2014), 여러 연구들을 통하여 지속적인 발전이 이루어지고 있다. R-CNN은 크게 Selective Search (Uijlings et al., 2013)를 사용하는 지역 제안(Region proposal)과 CNN을 이용한 특징 추출 및 분류를 시행하는 2가지 단계로 진행된다. 하지만 R-CNN 또한 높은 연산 능력을 요구하는 문제를 가지므로 향상된 연산 속도를 이룬 Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015), R-FCN (Dai et al., 2016), SSD (Liu et al., 2016), YOLO (Redmon et al., 2016)와 같은 발전된 알고리즘들이 제안되었고, 제안된 물체 탐지 기반 알고리즘들을 안전모 탐지에 적용하는 연구가 다양하게 진행되고 있다. Fang et al.(2018)는 Faster R-CNN을 사용하여 안전모를 쓰지 않고 있는 사람을 탐지했다. 하지만 학습 데이터로 건설현장 감시 동영상 데이터에서 추출한 10만 장이 넘는 현장 노동자의 영상을 사용했기 때문에 현실적으로 이러한 양의 데이터 확보가 어려울 수가 있고 학습에 오랜 시간이 걸린다는 문제점이 있다.

적은 양의 데이터를 사용하여 효율적으로 물체 탐지 알고리즘을 학습하기 위해 미리 학습된 알고리즘을 활용하는 Transfer Learning 기법이 여러 물체 탐지 연구에서 진행되었다. Mistry et al.(2017)는 미리 학습된 YOLO의 개량형인 YOLO2 (Redmon and Farhadi, 2017)를 이용해 오토바이 탑승자의 헬멧 착용 여부를 탐지하는 시스템을 구축하였고 Kim et al.(2017)은 미리 학습된 R-FCN을 사용해서 건설장비들을 탐지하여 건설현장을 효율적으로 관리하는 방법을 제시하였다. 사용된 학습 데이터는 각각 약 4천, 3천 장으로 학습에 필요한 영상의 양을 획기적으로 감소시켰다.

1.3 연구 목적

해외 연구들이 딥러닝 기반의 물체 탐지 알고리즘으로 건설현장에서 안전모를 탐지하는 방법이 효과적임을 증명하였지만, 아직까지 국내에선 원격 센서 및 HOG와 SVM 기반의 연구 및 기술이 개발된 사례가 몇몇 존재할 뿐 딥러닝 기반의 탐지와 관련된 연구 및 기술개발 사례가 전무하다. 따라서 본 연구에서는 딥러닝 기반의 물체 탐지 알고리즘 중 하나인 R-FCN을 사용하여 건설현장 내의 안전모를 탐지하는 기술의 활용을 제안하고자 한다. 특히 적은

양의 데이터로도 효과적으로 성능을 낼 수 있도록 Transfer Learning 기법을 도입하였다. 본 연구가 가지는 의의는 다음과 같다. 첫째, 국내 건설안전 분야에 딥러닝 기술의 도입을 제안하여 건설현장의 안전관리 향상의 가능성을 보인다. 둘째, Transfer Learning 기법의 적용을 통하여 기존의 건설 안전모 탐지처럼 학습에 많은 양의 데이터를 사용하지 않고 적은 양의 학습 데이터로도 건설 안전모를 효율적으로 탐지할 수 있다는 사실을 입증한다.

2. 연구 방법

본 연구에서는 영상 데이터를 기반으로 물체 탐지 알고리즘 중 하나인 R-FCN을 Transfer learning 기법을 통해 사용하여 안전모를 탐지하는 방법을 제안한다. R-FCN 알고리즘과 Transfer learning 기법에 대해서는 각각 2.1장과 2.2장에서 자세히 설명하겠다.

2.1 Region-based Fully Convolutional Network

선행 연구에서는 물체 탐지 알고리즘으로 Faster R-CNN과 YOLO2를 사용하여 안전모를 탐지하였지만 본 연구에서는 다음과 같은 이유로 R-FCN을 사용하였다. 개선된 Faster R-CNN 알고리즘인 Faster R-CNN++의 경우, PASCAL VOC 테스트 데이터셋을 기준으로 mAP가 83.8 %으로 82.0 %인 R-FCN보다 정확도는 1.8 %로 높지만 이미지 당 물체 인지 속도는 Faster R-CNN++은 3.36초이고 R-FCN은 0.17초로 나타났다(Dai et al., 2016). 정확도 1.8 %의 향상을 위해서 인지 속도가 약 20배 느린 Faster R-CNN++을 쓰는 것은 효율적이지 않다고 판단된다. YOLO2의 경우, 첫째로 COCO 데이터셋을 사용한 성능 측정 결과를 살펴보면 YOLO2보다 R-FCN이 성능이 더 높게 평가되었다(Dai et al., 2016; Redmon and Farhadi, 2017). 두번째로 YOLO2의 경우,

학습에 사용된 영상 데이터의 해상도에 따라 성능이 달라지는 경향을 보인다(Redmon and Farhadi, 2017). 따라서 더 안정적이고 좋은 성능을 보이는 알고리즘인 R-FCN을 사용하였다.

기존의 R-CNN 계열의 물체 탐지 알고리즘은 2가지 단계로 진행이 된다. 이러한 2가지 단계 진행의 가장 큰 단점은 매우 높은 계산량을 요구한다는 것이다. 그 이유는 1단계에서 제안되는 수 천개의 예상 지역에 대하여 반복적으로 계산이 수행되기 때문이다. R-FCN에서는 이러한 문제를 해결하기 위해서 전체 영상 내의 모든 계산을 공유하는 Fully Convolutional Operation 방식을 적용했고 이를 위해 Position-Sensitive Score Map이라는 개념을 제시했다. 이러한 변화를 통해서 R-FCN은 기존의 R-CNN보다 연산속도가 수십 배 빠르면서 탐지 성능은 다른 최첨단 기술에 못지 않은 결과를 보였다(Dai et al., 2016).

2.1.1 Region Proposal Network

R-FCN에선 R-CNN과 같이 대중적인 2단계의 물체 탐지 전략을 사용한다. 그중 첫 번째 단계인 후보 지역을 제안하는 단계에선 Region Proposal Network (RPN) (Ren et al., 2015)를 사용했다. 이때 RPN에서 사용되는 Feature들은 R-FCN의 메인 과정에서도 공유되므로 Weight Layer의 연산이 획기적으로 줄어들게 된다. 이렇게 제안된 관심영역(Region of Interest, RoI)들은 R-FCN의 메인 구조에서 물체 또는 배경으로 분류가 된다(Fig. 1).

2.1.2 ResNets

R-FCN의 근간이 되는 CNN 구조는 ResNets이다. ResNets의 가장 큰 특징은 회선층의 입력 자체를 회선층의 출력에 그대로 연결시키고(Skip Connection) 회선층의 결과에 더한 다음 활성화 함수 Relu (Maas et al., 2013)를 적용시켰다는 점이다. 이로 인해 딥러닝 네트워크의 고질적인 문제인 기울기 값의 소실 문제

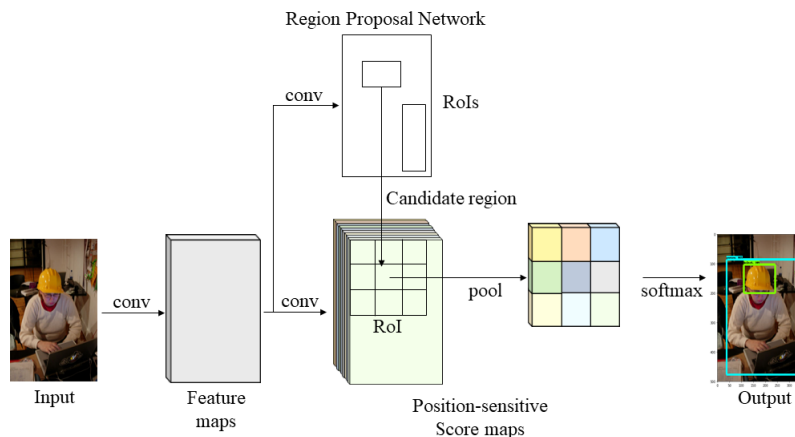


Fig. 1. Architecture of R-FCN

(Gradient Vanishing Problem)를 해결할 수 있게 되어 숨김 계층을 매우 많이 사용 할수 있게 되었다(He et al., 2016). R-FCN에선 ResNets 구조에서 Average Pooling과 완전 연결층을 제거해서 특징 맵만을 연산할 수 있도록 하였다.

2.1.3 Position-Sensitive Score Maps

기존의 네트워크들의 출력(Output) 형태와 달리 R-FCN에서는 Heat map 형태인 Position-Sensitive Score Maps을 출력으로 한다. 특정 클래스의 물체가 있는 경우 높은 값을 갖도록 클래스 별 Heat map을 설계하고 그 Heat map들을 RPN에서 제안한 RoI에 따라 잘라낸 다음 쌓아서 일정한 수의 그리드(Grid)를 갖는 Tensor로 출력을 나타낸다. 이때 그리드의 수는 Heat map내의 위치와 연관이 있는데 예를 들어 3x3 Tensor는 왼쪽위, 가운데위, 오른쪽위 등 각각의 영역을 그리드가 나타내는 것이 된다. 따라서 Tensor, 즉 Score map은 위치에 민감하게(Position-Sensitive) 된다. 이렇게 얻어진 Score Map은 평균화를 통해 해당 클래스에 대한 신뢰(Confidence) 정보로 변환되고 각 클래스 별 신뢰 정보는 Softmax 함수(Nasrabadi, 2007)를 통해 분류를 수행한다(Fig. 1).

2.2 Transfer Learning

본 연구에서 사용한 영상 데이터셋의 양이 약 1000장 정도 밖에 되지 않기 때문에 Transfer Learning 기법을 적용하여 약 33만장의 영상과 이를 분류하는 80개의 카테고리로 구성된 COCO

데이터셋(Lin et al., 2014)으로 미리 학습된 Resnet101 (He et al., 2016)을 R-FCN에 도입하였다. 딥러닝 네트워크의 경우, 구조 자체가 매우 복잡하기 때문에 학습 데이터가 적을수록 과최적화(Overfitting) 등과 같은 문제를 일으켜 성능이 떨어지는 반면에, 학습에 사용되는 데이터 양이 많을수록 딥러닝 네트워크의 성능이 극대화된다는 것은 여러 연구를 통하여 증명된 사실이다(Liu et al., 2017). 하지만 현실에서 특정 분야에 대한 많은 양의 데이터를 수집하기는 어려움이 따르기 때문에 비슷한 분야의 풍부한 데이터로 미리 학습된(Pre-trained) 모델로 특정 분야의 데이터를 학습하는 Transfer Learning 기법이 다양한 딥러닝 응용분야에서 많이 사용되고 있다(Weiss et al, 2016). Transfer learning에 대한 자세한 설명은 Shao et al.(2015)에 소개되어 있다. Fig. 2는 기존의 학습 방법과 Transfer learning 방법의 차이에 대한 기본적인 개념도이다.

3. 실험

안전모 자동 탐지 시스템의 개발 목적은 건설현장에서 건설노동자의 안전장비 착용 소홀을 방지하는 것에 있다. 기본적으로 안전관리자가 건설현장을 순찰하지만, 안전관리자가 건설현장이 아닌 장소에서도 영상 데이터를 통해 건설노동자의 안전장비 착용 여부를 확인할 수 있도록 한다.

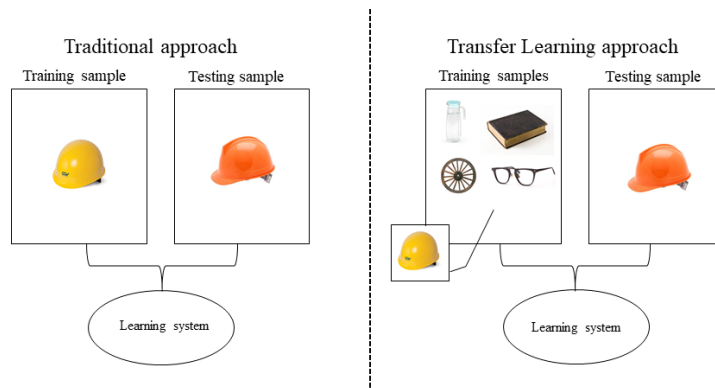


Fig. 2. Basic Frameworks of Traditional Approach and Transfer Learning Approach



Fig. 3. Workflow of the Experiment

3.1 실험 절차

본 연구에서는 크게 3단계로 실험을 진행하였다(Fig. 3). 첫 번째 단계는 데이터셋 구축으로 안전모 탐지 모델의 학습 및 평가를 위한 데이터를 오픈소스 영상 웹사이트에서 수집하였다(3.3 데이터셋 부분 참조). 두 번째 단계는 구축된 데이터셋을 COCO 데이터셋으로 미리 학습된 R-FCN 알고리즘(2.2 Transfer Learning 부분 참조)으로 학습하고 학습이 완료된 후, 평가를 수행한다. 세 번째 단계는 실제 촬영한 영상을 통한 탐지 모델의 성능 분석이고, 이를 좀 더 체계적으로 수행하기 위하여 여러 가지 상황들을 가정하여 안전모 탐지에 대한 분석을 수행하였다.

3.2 실험 환경

본 연구에서 사용한 하드웨어 사양은 Intel i7-4470 하스웰, Nvidia GeForce GTX 1080 그리고 RAM 32GB이다. 주요 소프트웨어 환경은 Ubuntu 16.04, Python 3.6 그리고 Tensorflow 1.3 (Abadi et al., 2016)이다. 사용한 R-FCN 알고리즘의 코드는 Tensorflow 라이브러리에서 도입했다.

3.3 데이터셋

학습 데이터는 건설노동자가 착용한 안전모만을 대상으로 한 것이 아니라 작업현장 내의 건설노동자들도 또한 대상으로 수집하였다. 그 이유는 건설노동자들이 안전모를 제대로 착용하고 있지 않을 가능성이 높다는 것을 경고하기 위한 의도이다.

R-FCN 알고리즘의 학습을 위한 영상은 연구, 개발용의 오픈소스 영상 웹사이트인 ImageNet (Deng et al., 2009)에서 Safety hat이라는 키워드로 학습용 영상을 탐색했다. 그중 현재 연결되지 않는 URL 및 불량 영상을 필터링하고 웹 검색엔진에서 저작권이 없는 안전모 관련 영상을 추가로 수집하여 총 1089장의 영상을 학습용 영상 데이터셋으로 구축하였다. 학습된 모델의 정성적 분석을 위하여 실제로 영상 159장을 촬영하여 사용하였다. 이중 건설현장 내의 여러 약조건 환경에서의 안전모 탐지에 대한 분석을 시행하기 위하여 4가지 상황(원거리의 촬영, 명암의 고저, 건설노동자의 겹침, 특이한 자세)으로 100장을 분류하였다(Table 1). 건설현장에서 촬영한 영상을 확보하는 것이 이상적이지만, 기술적 어려움으로 인하여 현장과 비슷한 환경에서 촬영한 사진을 사용하였다.

구축된 데이터셋을 약 4:1의 비율로 무작위로 나눈 후 이들을

각각 학습을 위한 데이터와 학습된 모델의 평가를 위한 데이터로 분리하였다. 그 후, 동일한 방법으로 5번의 교차검증을 수행하였다.

3.4 평가 방법

실험의 평가를 위한 기준으로는 mean Average Precision (mAP)을 사용하였다. mAP는 PASCAL VOC에서 평가 기준으로 사용된 지표로서, 각 분류 클래스별 Average Precision (AP)의 평균으로 물체탐지 알고리즘의 성능을 한가지 지표로 나타낼 수 있다(Everingham et al., 2010). 일반적으로 탐지 모델의 성능을 평가하는데 사용되는 지표는 정밀도(precision)와 검출율(recall)이다. 정밀도는 Eq. (1)과 같이 물체를 탐지한 전체 경우에서 참값을 탐지한 경우에 대한 비율을 뜻하고 검출율은 Eq. (2)와 같이 옳게 탐지한 경우에서 참값을 탐지한 경우에 대한 비율을 뜻한다. 이 두 가지 지표는 일반적으로 서로 음의 상관관계를 가지기 때문에 정밀도-검출율 그래프에서 그래프 아래의 면적으로 정의되는 AP를 사용하며, 이때 AP가 1에 가까울수록 물체탐지 알고리즘의 성능이 높다고 할 수 있다.

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (1)$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (2)$$

4. 결과 분석

학습이 완료된 모델에 대하여 성능 평가를 실시한 결과, 사람(건설노동자)과 안전모의 평균 mAP값이 0.85로 계산되며, 분류 클래스 별 mAP는 안전모는 0.83, 사람은 0.86으로 평가되었다(Table 2). 본 연구에서 사용한 물체탐지 알고리즘인 R-FCN의 기준 mAP가 0.83인 것을 감안하면(Dai et al., 2016) 학습이 정상적으로 진행되었음을 알 수 있다. 안전모에 대한 mAP보다 사람에 대한 mAP가 더 높게 측정된 이유는, Transfer learning으로 사용된 데이터인 COCO 데이터셋의 분류 클래스에 사람은 있지만 안전모는 없기 때문으로 보여진다. 본 연구에서는 COCO 데이터셋으로 미리 학습된 ResNets 알고리즘이 R-FCN에 적용되었으므로 안전모의 경우, 학습에 사용된 데이터가 ImageNet에서 수집한 영상들로부터 이루어

Table 1. Number of Images for Each Environment

Category	Number of images
Long distance	25
Brightness	25
Overlap	25
Extraordinary pose	25

Table 2. Average mAP for Each Class

Class	mAP
Safety Hat	0.8309
Person	0.8623
Average	0.8466
Dai et al., 2016	0.83

진 반면에 사람의 경우, ImageNet에서 수집된 데이터뿐만 아니라 Transfer learning에 사용된 데이터도 영향을 받았다.

영상 데이터셋 중 평가용 데이터로 직접 영상 내에서 사람과 안전모가 탐지되는 모습을 확인해 본 결과, 대부분 Fig. 4와 같이 정상적으로 탐지되는 것을 확인할 수 있었다. 오탐지 된 경우는

다음 같은 특징을 가지고 있었다. 첫째로 원거리에서 촬영된 경우 (Fig. 5(a)), 두 번째로는 사람들이 겹쳐 있는 경우(Fig. 5(b)), 세 번째로 명암이 너무 높거나 낮을 경우(Fig. 5(c)), 마지막으로 사람들이 특이한 자세를 취하는 경우이다(Fig. 5(d)).

제안 방법의 성능을 정성적으로 분석하기 위해 ImageNet 데이

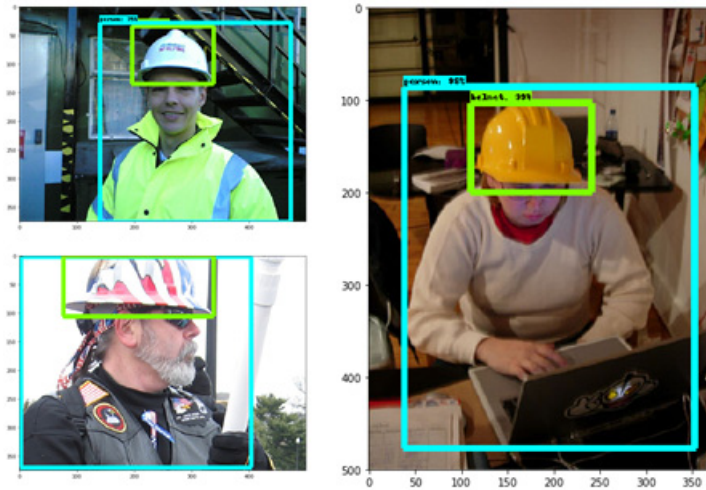


Fig. 4. Successful Detections with ImageNet

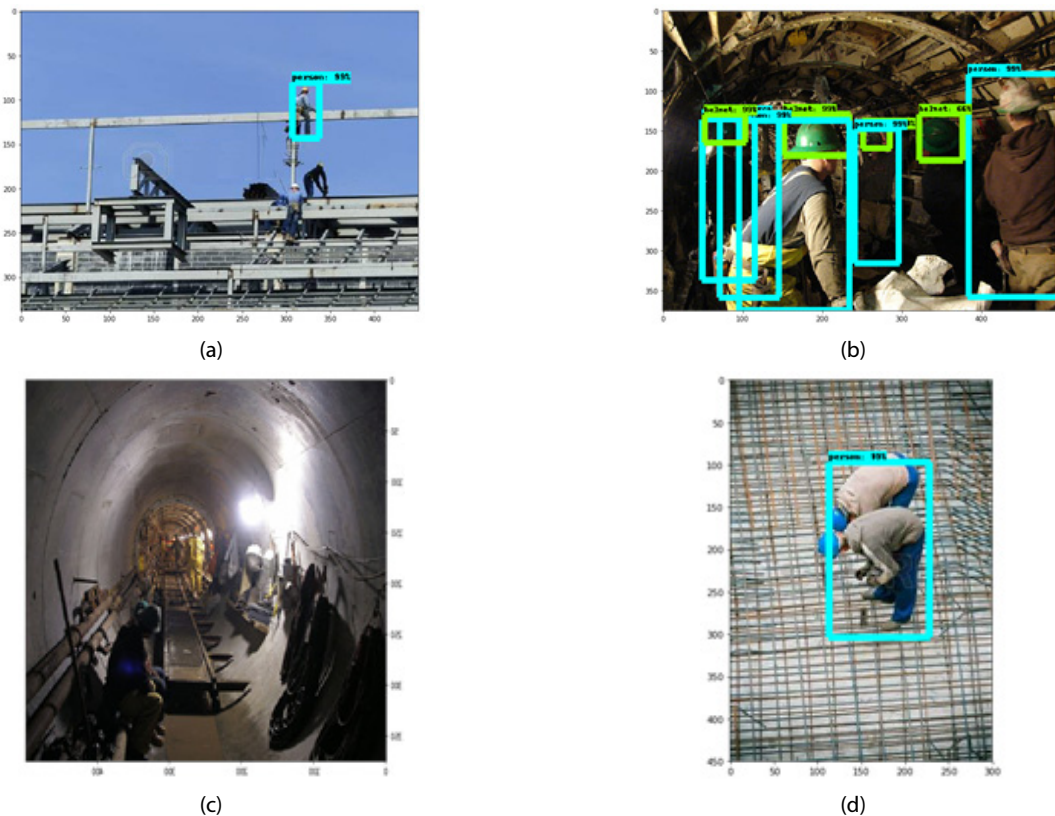


Fig. 5. Cases of Detection Fail with ImageNet

Table 3. mAP for Each Class in Different Situations

Category	Helmet	Person	Average
Long distance	0.3319	0.6473	0.4896
Brightness	0.7038	0.8058	0.7548
Overlap	0.7571	0.7765	0.7668
Extraordinary pose	0.7340	0.7575	0.7457



Fig. 6. Successful Detections with Actual Photos

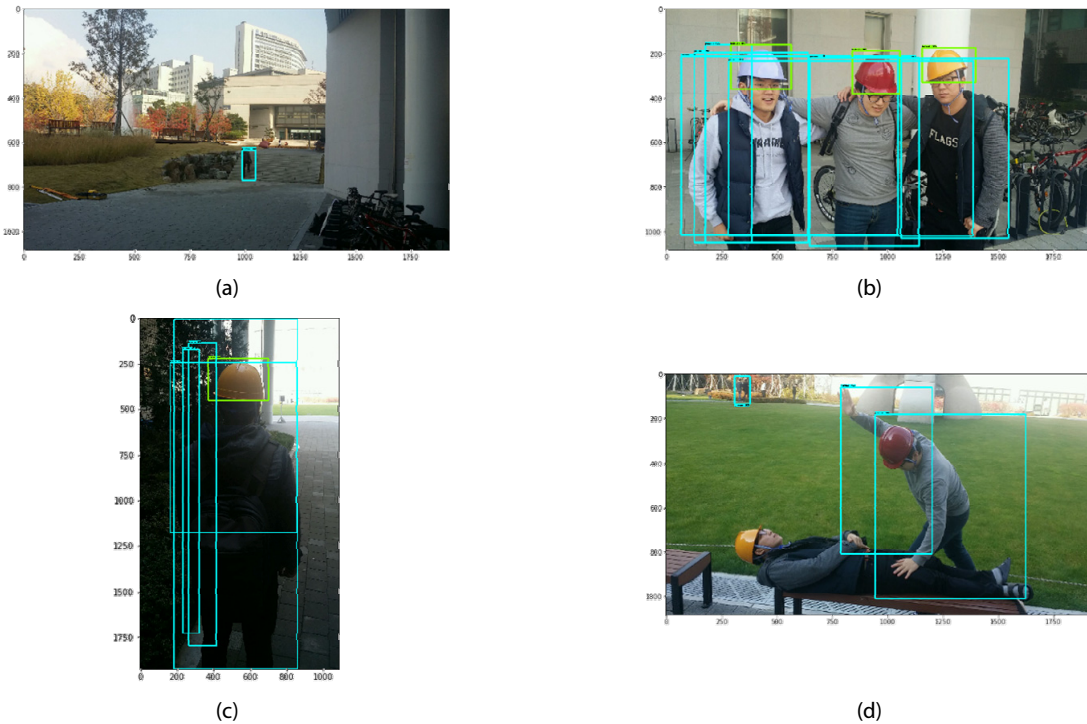


Fig. 7. Cases of Detection Fail with Actual Photos

터가 아닌 직접 촬영한 사진 데이터로 안전모 탐지를 시행했다. 일반적인 촬영 영상은 Fig. 6과 같이 정상적으로 안전모를 탐지하는 결과를 확인하였다. 추가로, ImageNet 데이터로 오탐지한 경우가

많았던 4가지 약조건의 상황을 가정하고 촬영한 영상에 대한 분석을 수행하였다. 4가지 약조건 상황에 따른 탐지 평가 결과는 Table 3과 같다. 약조건 상황 중에서 Fig. 7(a)와 같이 촬영 거리가 원거리인

경우가 mAP가 가장 나쁘게 평가되었다. 그 원인은 알고리즘 구조상의 문제 때문으로 판단되는데, 영상이 Pooling 단계를 거칠수록 차원의 크기는 줄어들기 때문에 애초부터 적은 수의 픽셀들로 구성된 원거리 대상은 전역 특징이 제대로 추출되지 않기 때문이다. 특히 안전모의 경우, 사람보다 전역 특징이 단순하여 이러한 영향을 많이 받았을 것이라 판단된다. Fig. 7(b)와 같이 여러 사람이 겹쳐져서 촬영된 경우는 사람 대한 탐지가 정확하게 하나로 이루어지지 않고 여러 개로 탐지되는 현상을 확인할 수 있다. 사람의 경우, 영상 내에서 전체 모습이 나타나는 경우뿐만 아니라 상반신과 같이 몸의 일부분만 나타나는 경우가 많다. 따라서 이러한 몸의 일부분과 몸 전체가 동일하게 학습되기 때문에 영상 하나에 여러 사람이 겹쳐져 있으면, 탐지기가 겹쳐진 몸의 부분을 사람으로 인식하여 여러 사람으로 탐지하는 결과를 낳은 것으로 판단된다. Fig. 7(c)와 같이 주변 환경의 명암이 뚜렷하지 못하거나 Fig. 7(d)처럼 사람이 서 있지 않고 몸을 숙이거나 엎드린 자세와 같이 특수한 자세를 취하면 잘 탐지하지 못하는 결과를 보였다. 특수한 상황에서 물체 탐지 능력이 떨어지는 이유는 부족한 학습 데이터의 양과 질에 있다고 추정되는데, 학습에 사용된 ImageNet의 데이터가 대부분 명암이 뚜렷하고 서 있는 자세와 같은 보편적인 사람의 모습으로 구성되어 있다. 따라서 학습을 많이 하지 못한 특수한 환경에 대해서는 탐지율이 떨어지는 것으로 판단된다.

5. 결론

본 연구는 건설현장에서 효율적인 안전 관리를 위해 영상 데이터를 기반으로 R-FCN 물체 탐지 알고리즘을 이용한 건설 안전모 자동탐지 시스템 구축을 제안하였다. 시스템 구축에 사용된 데이터는 ImageNet에서 획득한 사람 및 안전모가 포함된 총 1089개의 영상 데이터이다. 1000여장 규모의 데이터로는 딥러닝 학습을 수행하기에 데이터의 양이 부족하므로 미리 학습된 ResNets 알고리즘을 사용하는 Transfer Learning 기법을 적용하였다.

학습된 모델을 평가해본 결과 탐지 성능을 나타내는 지표인 mAP가 평균 0.85로 평가되었다. 이는 R-FCN의 성능 mAP인 0.83보다 높은 값이므로 모델이 효과적으로 학습되었음을 보여주며, 적은 양의 데이터로도 Transfer Learning을 통해 기존의 많은 양의 데이터로 학습한 안전모 탐지 모델에 뒤처지지 않는 성능을 낼 수 있음을 확인하였다. 따라서 해외 건설현장보다 관련 데이터가 부족한 국내 건설현장의 상황에서도 본 연구의 안전모 탐지 모델이 충분히 도입 가능하다는 것을 사실을 입증한다.

실제 촬영한 영상 데이터로 오탐지 사례에 대한 분석을 시행해본 결과, 특수한 상황에서는 잘 탐지가 되지 못한다는 한계가 발견되었다. 그 이유는 크게 알고리즘의 구조적인 한계와 풍부하지 못한 학습 데이터의 질과 양이라고 판단된다. 따라서 이러한 문제점을

보완하기 위하여 다음과 같은 추가 연구를 진행할 예정이다. 첫 번째로 영상 데이터의 추가 수집 및 학습 후, 다시 탐지 테스트를 진행할 예정이다. 현장 여건상, 원거리에서 습득한 데이터를 중심으로 학습하는 것이 무엇보다 중요한 것으로 판단되지만 본 연구에서는 데이터 습득의 기술적 한계로 인하여 확보를 하지 못하였다. 두 번째로는 구조적 한계점으로 인한 문제점을 해결하기 위하여 더 발전된 알고리즘을 적용해 보는 것이다. 새로운 CNN (Zhu and Newsam, 2017; Liu and Zeng, 2018) 및 물체 탐지 알고리즘 (Lin et al., 2017; Li et al., 2018)이 지금도 계속 발표되고 있다. 이러한 새로운 알고리즘을 적용 후 테스트를 시행하여 더 좋은 결과가 나오는지 확인할 예정이다. 마지막으로 시스템적인 보완 사항으로 같은 탐지 화면 내의 건설노동자와 안전모의 수가 일치하지 않을 경우, 건설노동자들이 안전모를 정상적으로 착용하지 않을 가능성이 높은 상황으로 간주하고 관리자에게 자동으로 경고를 하는 기능을 추가할 계획이다.

본 연구에서 제안한 영상 기반의 안전모 탐지 시스템이 실제 건설현장에서 건설노동자들의 안전성 향상에 기여 하는 것뿐만 아니라 안전모 탐지를 넘어서서 다양한 건설 분야에서의 물체 탐지에 활용될 수 있을 것으로 기대된다.

감사의글

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2018R1A2B2009160).

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. and Zheng, X. (2016). "Tensorflow: a system for large-scale machine learning." *USENIX Symposium on Operating Systems Design and Implementation*, Vol. 16, pp. 265-283.
- Dai, J., Li, Y., He, K. and Sun, J. (2016). "R-fcn: Object detection via region-based fully convolutional networks." *Advances in Neural Information Processing Systems*, pp. 379-387.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K. and Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database." *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 248-255.
- Everingham, M., Van Gool, L., Williams, C. K., Winn, J. and Zisserman, A. (2010). "The pascal visual object classes (voc) challenge." *International Journal of Computer Vision*, Vol. 88, No. 2, pp. 303-338.

- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M. and An, W. (2018). "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos." *Automation in Construction*, Vol. 85, pp. 1-9.
- Girshick, R. (2015). "Fast r-cnn." *Proc. of the IEEE International Conference on Computer Vision*, IEEE, pp. 1440-1448.
- Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 580-587.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). "Deep residual learning for image recognition." *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 770-778.
- Illingworth, J. and Kittler, J. (1988). "A survey of the hough transform." *Computer Vision, Graphics, and Image Processing*, Vol. 44, No. 1, pp. 87-116.
- Kim, H., Kim, H., Hong, Y. W. and Byun, H. (2017). "Detecting construction equipment using a region-based fully convolutional network and transfer learning." *Journal of Computing in Civil Engineering*, ASCE, Vol. 32, No. 2, pp. 04017082.
- Li, Z., Chen, Y., Yu, G. and Deng, Y. (2018). "R-FCN++: Towards accurate region-based fully convolutional networks for object detection." *Proc. of the Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans.
- Lin, T. Y., Dollár, P., Girshick, R. B., He, K., Hariharan, B. and Belongie, S. J. (2017). "Feature pyramid networks for object detection." *CVPR*, Vol. 1, No. 2, p. 4.
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L. (2014). "Microsoft coco: Common objects in context." *Proc. of European Conference on Computer Vision*, Springer, Cham, pp. 740-755.
- Liu, B., Wei, Y., Zhang, Y. and Yang, Q. (2017). "Deep neural networks for high dimension, low sample size data." *Proc. of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, Melbourne, pp. 2287-2293.
- Liu, W. and Zeng, K. (2018). "SparseNet: A sparse densenet for image classification." *arXiv preprint arXiv:1804.05340*.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. and Berg, A. C. (2016). "Ssd: Single shot multibox detector." *Proc. of European Conference on Computer Vision*, Springer, Cham, pp. 21-37.
- Maas, A. L., Hannun, A. Y. and Ng, A. Y. (2013). "Rectifier nonlinearities improve neural network acoustic models." *Proc. of ICML*, Vol. 30, No. 1, p. 3.
- Memarzadeh, M., Golparvar-Fard, M. and Niebles, J. C. (2013). "Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors." *Automation in Construction*, Vol. 32, pp. 24-37.
- Mistry, J., Misra, A. K., Agarwal, M., Vyas, A., Chudasama, V. M. and Upla, K. P. (2017). "An automatic detection of helmeted and non-helmeted motorcyclist with license plate extraction using convolutional neural network." *Proc. of 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, pp. 1-6.
- Nasrabadi, N. M. (2007). "Pattern recognition and machine learning." *Journal of electronic imaging*, Vol. 16, No. 4, p. 049901.
- Occupational Safety and Health Research Institute (OSHRI) (2016). *Cause of Industrial Accident in 2014*, OSHRI Research Report (in Korean).
- Redmon, J. and Farhadi, A. (2017). "YOLO9000: better, faster, stronger." *arXiv preprint*.
- Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). "You only look once: Unified, real-time object detection." *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 779-788.
- Ren, S., He, K., Girshick, R. and Sun, J. (2015). "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in Neural Information Processing Systems*, pp. 91-99.
- Rubaiyat, A. H., Toma, T. T., Kalantari-Khandani, M., Rahman, S. A., Chen, L., Ye, Y. and Pan, C. S. (2016). "Automatic detection of helmet uses for construction safety." *Proc. of 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW)*, IEEE, pp. 135-142.
- Shao, L., Zhu, F. and Li, X. (2015). "Transfer learning for visual categorization: A survey." *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, No. 5, pp. 1019-1034.
- Silva, R., Aires, K., Santos, T., Abdala, K., Veras, R. and Soares, A. (2013). "Automatic detection of motorcyclists without helmet." *Proc. of Computing Conference (CLEI), 2013 XXXIX Latin American*, IEEE, pp. 1-7.
- Uijlings, J. R., Van De Sande, K. E., Gevers, T. and Smeulders, A. W. (2013). "Selective search for object recognition." *International Journal of Computer Vision*, Vol. 104, No. 2, pp. 154-171.
- Vishnu, C., Singh, D., Mohan, C. K. and Babu, S. (2017). "Detection of motorcyclists without helmet in videos using convolutional neural network." *Proc. of 2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp. 3036-3041.
- Weiss, K., Khoshgoftaar, T. M. and Wang, D. (2016). "A survey of transfer learning." *Journal of Big Data*, Vol. 3, No. 1, p. 9.
- Wen, C. Y., Chiu, S. H., Liaw, J. J. and Lu, C. P. (2003). "The safety helmet detection for ATM's surveillance system via the modified Hough transform." *Proc. of the IEEE 37th Annual 2003 International Carnahan Conference on Security Technology*, IEEE, pp. 364-369.
- You, H. J., You, Y. T. and Kang, K. S. (2017). "A study of the efficiency improvement of the safety management personnel system in apartment construction site." *Korea Safety Management & Science Korea Safety Management & Science*, Vol. 19, No. 1, pp. 87-94 (in Korean).
- Yuen, H., Princen, J., Illingworth, J. and Kittler, J. (1990). "Comparative study of Hough transform methods for circle finding." *Image and Vision Computing*, Vol. 8, No. 1, pp. 71-77.
- Zhu, Y. and Newsam, S. (2017). "Densenet for dense flow." *Proc. of 2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 790-794.