ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2018 March: 24(1): 1~23 http://dx.doi.org/10.13088/jiis.2018.24.1.001

온라인 쇼핑몰에서 상품 설명 이미지 내의 키워드 인식을 위한 딥러닝 훈련 데이터 자동 생성 방안*

김기태
㈜마인드그룹 개발본부
(*ktkim@mindgroup.co.kr*)

오원석
한양대학교 경영학부
(*ows090@hanyang.ac.kr*)

임근원
한양대학교 경영학부
(*wom2277@hanyang.ac.kr*)

차은우
한양대학교 경영학부
(*eunwoo921@hanyang.ac.kr*)

신민영
한양대학교 중어중문학과
(*minhayang@hanyang.ac.kr*)

김종우
한양대학교 경영대학
(*kjw@hanyang.ac.kr*)

E-commerce 환경의 발전으로 소비자들은 다양한 상품들을 한 자리에서 폭 넓게 비교할 수 있게 되었다. 하지만 온라인 쇼핑몰에 올라와있는 상당량의 주요 상품 정보들이 이미지 형태이기 때문에 컴퓨터가 인지할 수 있는 텍스트 기반 검색 시스템에 반영될 수 없다는 한계가 존재한다. 이러한 한계점은 일반적으로 기존 기계학습 기술 및 OCR(Optical Character Recognition) 기술을 활용해, 이미지 형태로 된 키워드를 인식함으로써 개선할 수 있다. 그러나 기존 OCR 기술은 이미지 안에 글자가 아닌 그림이 많고 글자 크기가 작으면 낮은 인식률을 보인다는 문제가 있다. 이에 본 연구에서는 기존 기술들의 한계점을 해결하기 위하여, 딥러닝 기반 사물인식 모형 중 하나인 SSD(Single Shot MultiBox Detector)를 개조하여 이미지 형태의 상품 카탈로그 내의 텍스트 인식모형을 설계하였다. 하지만 이를 학습시키기 위한 데이터를 구축하는 데 상당한 시간과 비용이 필요했는데, 이는 지도학습의 방법론을 따르는 SSD 모형은 훈련 데이터마다 직접 정답 라벨링을 해줘야 하기 때문이다. 본 연구는 이러한 문제점을 해결하기 위해 '훈련 데이터 자동 생성 프로그램'을 함께 개발하였다. 훈련 데이터 자동 생성 프로그램을 통해 수작업으로 데이터를 만드는 것에 비하여 시간과 비용을 대폭 절감할 수 있었으며, 생성된 훈련용 데이터를 통해 모형의 인식 성능을 높일 수 있었다. 더 나아가 실험연구를 통해 자동으로 생성된 훈련 데이터의 특징별로 인식기 모형의 성능에 얼마나 큰 영향을 끼치는지 알아보고, 성능 향상에 효과적인 데이터의 특징을 분석하였다. 본 연구를 통해서 개발된 상품 카탈로그 내 텍스트 인식모형과 훈련 데이터 자동 생성 프로그램은 온라인 쇼핑몰 판매자들의 상품 정보 등록 수고를 줄여줄 수 있으며, 구매자들의 상품 검색 시 결과의 정확성을 향상시키는 데 기여할 수 있을 것으로 기대한다.

주제어 : 딥러닝, 훈련데이터 생성, OCR, 속성 기반 검색, Single Shot MultiBox Detector

논문접수일 : 2017년 7월 21일 논문수정일 : 2017년 12월 18일 게재확정일 : 2018년 1월 14일
원고유형 : 일반논문(급행) 교신저자 : 김종우

1. 서론

21세기 이후 인터넷과 정보통신기술이 발전하면서 이를 활용한 다양한 비즈니스들이 등장하였는데, 이 중 아마존이나 옥션으로 대표되는 전자상거래 시장은 폭발적인 규모로 성장해 가고

* 이 논문은 '비즈니스 랩 기반의 빅인텔리전스 경영교육 사업단(CK2)'의 지원을 받아서 수행된 연구임

있다. 이는 국내 시장 역시 마찬가지로, 전자상거래 시장은 현재 전 세계 10위권 내에 위치할 정도이며, B2C 시장의 규모는 2010년에 27조 3000억 원을, 2016년에 53조 원대를 기록하며 큰 폭으로 성장 중이다(Yang, 2017). 이러한 성장 폭에서 볼 수 있듯이 전자상거래를 사용한 상품 구매는 일상생활 속에서 점차 당연한 현상이 되어가고 있으며, 이러한 소비자의 변화에 부응하듯 다양한 온라인 쇼핑몰들이 개점을 하고 있다(Hwang et al., 2005; Yang et al., 2016).

이렇게 전자상거래 시장이 커지면서 다양한 상품들이 온라인 쇼핑몰에 등록되고 있고, 소비자들은 원하는 상품을 쉽게 구매할 수 있게 되었다. 그러나 등록되는 상품 수가 많아지면서 소비자들은 원하는 물건을 찾는 데 어려움을 겪게 되었다. 예를 들어 소비자가 원하는 상품명을 일반화된 키워드로 입력할 경우 검색 결과가 너무 많이 나오는 문제가 발생하고 있으며, 반대로 세밀한 속성 키워드를 입력할 경우 검색 결과가 잘 나오지 않는 문제가 있는 것이다. 이에 전자상거래 상에서 소비자들의 편의를 증가시키기 위해 원하는 상품을 쉽게 찾을 수 있도록 검색 시스템을 어떻게 개선할 것인가에 대한 여러 연구가 이루어지고 있다(Mo and Lee, 2009).

이러한 전자상거래 상품 검색 시의 어려움의 원인 중에 하나는 상품 판매자가 올리는 상품 카탈로그가 이미지 형태라는 것이다. 텍스트화 되지 않은 이미지 속 상품 정보들은 검색의 대상이 되지 못해 검색시스템에 활용할 수 없다. 이러한 문제점을 해결하기 위해 기존 OCR 기술들을 활용하여 상품의 상세 설명이 적혀 있는 이미지에서 키워드를 추출한 후 검색 정보로 사용할 수 있다. 하지만 키워드의 크기가 작거나 키워드의 서체가 일정치 않은 경우 OCR 기술의 인식 성능이 낮아 활용에 어려움이 있다. 이에 본 연구는 최근 2010년대 들어 이미지 인식 분야에서 좋은 성능을 보이고 있는 딥러닝 기법을 사용하여 이미지 형태의 상품 상세 설명 부분(이하 카탈로그 이미지)에 있는 키워드를 인식하는 방안을 제시하도록 한다. 인식에 사용한 모형은 사물 인식 부분에서 좋은 인식 성능을 보여주고 있는 Single Shot Multibox Detector(SSD)로, 이를 활용하여 다양한 특정 키워드들을 인식할 수 있도록 설계하였다(Liu et al., 2016). 그러나 지도 학습을 사용해야 하는 SSD 모형의 특성상, 정답이 태깅되어 있는 대량의 훈련용 데이터가 필요하다는 것이 문제가 된다. 훈련용 데이터를 확보하기 위해 사람이 카탈로그 이미지에서 검색에 사용하고자 하는 키워드들의 위치와 텍스트 정보를 일일이 입력하여 훈련용 정답 데이터를 확보할 수 있으나, 이 방법은 오랜 시간이 소요된다. 또한 간혹 사람이 실수로 인식할 키워드를 보지 못하여 해당 키워드가 누락되거나, 제작 시간을 줄이기 위해 사람을 추가로 고용할 경우 인건비가 추가로 발생할 수 있다. 게다가 특정 검색 키워드를 훈련시키고자 할 때 해당 키워드가 포함된 데이터를 수집하는 것 또한 어렵다.

이러한 문제를 극복하기 위해 본 연구에서는 기존의 카탈로그 이미지들이 대다수 컴퓨터 환경에서 제작되었다는 것에서 착안하여, 자동으로 훈련 데이터를 생성하는 프로그램을 개발하였다. 이 프로그램은 이미지 속에 여러 키워드와 각종 요소들을 카탈로그 이미지와 유사하게 그려냄과 동시에 키워드의 위치 정보와 텍스트 정보를 바탕으로 정답 데이터를 생성한다. 이러한 '훈련용 데이터 자동 생성 프로그램'을 사용하면 기존에 사람이 직접 작업하던 것보다 효율적으로 대량의 훈련용 데이터를 구할 수 있다. 하지

만 다량의 학습데이터를 생성하더라도 생성된 데이터의 특징에 따라 그 학습효과가 달라지기 마련이다. 따라서 본 연구에서는 이 '훈련용 데이터 자동 생성 프로그램'으로 만든 데이터의 특징들이 SSD의 키워드 인식 성능에 미치는 영향을 분석하기 위한 실험을 수행하였다. 본 연구는 이 실험을 통해 텍스트 인식모형에 효과적인 학습 데이터 생성 방안을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 전자상거래 검색시스템, 딥러닝, OCR 등에 대하여 살펴본다. 3장에서는 훈련 데이터 자동 생성 방안을 제시한다. 4장에서는 제시한 방법의 성능 평가를 위한 실험과 성능에 영향을 주는 데이터 특징을 파악하기 위한 실험 설계와 결과를 제시한다. 5장에서는 본 연구의 의의와 한계점을 제시하도록 한다.

2. 관련 연구

2.1 전자상거래 검색시스템의 개선

전자상거래에서 고객들이 원하는 상품을 보여주는 시스템의 종류를 크게 두 가지로 나누어 볼 수 있다. 우선 첫 번째로는 온라인 상점이 직접 고객들에게 물품을 추천해주는 추천 시스템을 들 수 있는데, 이는 고객의 활동과 기존 구매 내역 등을 분석해 향후 고객이 어떤 것을 구매할지 자동으로 예상하여 제공해 주는 시스템이다(Cho et al., 2015). 이러한 추천시스템은 여러 전자상거래 플랫폼 내에 도입이 되면서 다양한 연구들이 이루어 졌는데, 그 중 가장 많이 연구된 방식은 협업 필터링 기법을 활용한 상품 추천이다(Kim, 2008). 이 외에도 고객의 구매 후기를 토픽 모델링 기법으로 분석하여 추천 모형을 개발하거나(Cho et al., 2015), 데이터 마이닝의 유전자 알고리즘을 이용하여 상품을 추천하거나(Kim and Kim, 2005), 고객의 인터넷 사용 패턴을 통해 고객의 관심 이슈를 분석하여 추천 시스템에 활용하는 등(Choi et al., 2015)의 연구들이 존재한다.

두 번째로는 고객들이 찾으려는 물건을 정확하게 찾아주는 검색 시스템을 들 수 있다. 검색 시스템은 사용자에게서 검색 키워드를 받아 해당 키워드를 포함하는 상품들을 보여주는데, 소수의 키워드에서 고객이 정말 원하는 것이 무엇인지를 찾아내기 위하여 다양한 자연어 처리 방법을 활용한 연구들이 등장하게 되었다. 이의 예로는 개념 망을 활용하여 관련 검색 키워드를 사용자에게 추천하거나(Ma et al., 2015), 쇼핑몰에서 서로 다르게 입력한 속성 키워드 등을 추출하여 온톨로지 서버를 구축하여 검색 성능을 높인 사례 등이 있다(Hwang et al., 2005). 이 외에 오픈마켓 카테고리 검색 시스템을 의미적으로 개선한 사례 등도 있다(Hong et al., 2012).

본 연구에서는 이 두 종류의 시스템 중 고객들이 원하는 상품을 보다 쉽게 찾을 수 있도록 온라인 상점의 검색 시스템 부분을 개선해 보고자 한다. 그러나 고객이 입력한 키워드를 확장시키거나 유사한 의미를 찾는 것이 아니라, 검색이 되는 영역을 넓혀 보다 많은 키워드를 추출하려는 면에서 다른 기존 연구들과는 접근 방법이 다소 다르다고 볼 수 있다.

2.2 딥러닝

최근에 많은 관심을 받고 있는 딥러닝은 인공신경망에 그 기반을 두고 있으며, 그 역사를 거

슬러 올라가다보면 1950년대에 사람의 뇌를 모방하는 것을 목적으로 만들어진 '퍼셉트론(Perceptron)'을 그 시초로 볼 수 있다(Rosenblatt, 1958). 그러나 이 퍼셉트론은 간단한 XOR 문제도 풀 수 없다는 단층 신경망의 한계를 보이면서 한 차례의 겨울을 맞게 되었다(Minsky and Papert, 1969). 이러한 겨울 속에서 1980년대에 다층 신경망을 효과적으로 학습할 수 있는 역전파 알고리즘이 등장하게 되는데, 이를 통해 인공신경망 연구는 다시 한 번 부활할 수 있게 되었다(LeCun et al., 1989). 하지만 이 역시 여러 개의 은닉층을 가지는 신경망을 학습할 때 일어나는 '사라지는 경사(Vanishing gradient) 문제'가 등장하면서 성능을 끌어 올리는데 실패하게 되고, 1990년대에 들어서면서 결국 다시 한 번 겨울을 맞게 되었다. 이처럼 흔히 인공신경망이라는 분야로 연구되어 온 딥러닝은 과거 이론적인 문제점들과 함께 부족한 컴퓨팅 파워로 빛을 발하지 못하였다(Zhang, 2015).

인공신경망이 지니는 이러한 문제들은 2000년대에 들어서면서 컴퓨터의 성능이 비약적으로 향상되고, 사라지는 경사 문제를 해결하기 위한 여러 방법들이 연구자들에 의해 제시되면서 점차 해결되기 시작하였다. 또한 GPU(Graphic Process Units)의 많은 코어들을 활용하여 병렬처리를 기존보다 훨씬 빠르게 처리할 수 있게 되었으며, 학습에 필요한 데이터의 양도 인터넷의 발전과 함께 엄청난 속도로 쌓이기 시작하였다. 이러한 조건 하에서 딥러닝 연구는 점차 활기를 띠기 시작하였는데, 기존 연구들이 어려워하던 이미지 인식, 음성 인식 및 영상 인식 부분 등 다양한 분야에서 딥러닝 모형들이 강점을 보이며 활용되기 시작하였다. (Choi and Min, 2015; Kim, 2010; Zhang, 2015)

이와 같은 딥러닝의 여러 활용 분야들 중 이미지 인식 부분은 딥러닝이 특히 우수한 성능을 보이고 있다. 2012년 국제 이미지 인식 대회 중 하나인 ImageNet Challenge에서 생물의 시신경을 모방해 만든 CNN(Convolutional Neural Networks)을 쌓아 올려 만든 딥러닝 모형이 다른 기존 알고리즘에 비해 큰 성능 차이로 우승하면서 세간의 주목을 받기 시작하였는데, 이 대회 이후 각종 이미지 인식 분야에서 CNN을 필두로 한 딥러닝 모형들이 꾸준히 활약하게 되면서 이미지 관련 연구들의 흐름을 바꾸게 되었다(Kim et al., 2015; Krizhevsky et al., 2012).

이러한 딥러닝 기반 이미지 인식 연구는 사물 인식 연구에도 기여하였다. R-CNN(Regions with CNN features)은 기존 딥러닝 기반 이미지 인식 모형에 위치 검출 기능을 추가하여, 딥러닝 기법을 활용한 사물 인식 모형을 구현하였다(Girshick et al., 2014). 다만 이는 여러 연산 과정을 거쳐야 하기 때문에 속도가 느리다는 단점이 있었다. 이를 개선하기 위해 사전 훈련 과정을 간소화한 Fast R-CNN(Girshick, 2015)과, 개별적으로 후보 영역을 선출해야하는 문제를 해결한 Faster R-CNN 등이 연구되었다(Ren et al., 2015). 최근에는 더욱 빠른 인식 속도를 위해 단일 모형으로 위치 검출과 사물 인식이 가능한 End-To-End 사물 인식 모형이 주목받고 있다. 이의 대표적인 예로 SSD(Single Shot Multibox Detector) 및 YOLO(You Only Look Once)와 같은 딥러닝 모형 등이 존재하는데, 이들은 실시간에 가깝게 이미지 내 사물의 위치와 종류를 파악해낼 수 있다는 강점을 가지고 있다(Liu et al., 2016; Redmon et al., 2016).

이 중 본 연구에서 활용하고 있는 SSD 모형은 모형 구조 중 특징 추출 부분과 검출 부분을 자

유롭게 변경할 수 있어, 인식하고자 하는 대상들의 특징에 맞춰 보다 적합한 모형을 생성할 수 있다는 장점을 지니고 있다. 그 예로 크기가 작은 사물들을 주로 인식하기 위해 'Anchor Box' 혹은 'Default Box', 'Prior Box' 등으로 불리는 검출 부분의 인식 범위를 보다 세밀하게 줄이거나(Liu et al., 2016), 특징 추출 부분에 활용되는 이미지 인식 모형 중 하나인 VGG16 모형의 연결 구조를 변형하여 성능을 개선한 사례가 있다(Cao et al., 2017). 이 외에도 사물 인식 성능을 높이기 위하여 검출 부분의 합성곱(Convolution) 층에 역합성곱(Deconvolution) 층 및 예측 모듈을 추가한 DSSD 모형 등이 존재한다(Fu et al., 2017).

2.3 광학적 문자 인식

광학적 문자 인식(Optical Character Recognition; 이하 OCR)은 광학 메커니즘을 통해 디지털 이미지에 있는 이미지 형태의 텍스트를 편집할 수 있는 텍스트 형태로 변환시켜주는 프로세스를 칭한다(Singh, 2013). 이러한 OCR은 1870년 C. R. Carey가 광전지 모자이크를 사용하는 이미지 전송 시스템인 망막 스캐너를 발명하면서 시작되었고, 1940년대 중반 디지털 컴퓨터가 개발되면서 본격화 되었다(Eikvil, 1993). 초창기의 OCR 기술은 한정적인 글씨체만을 인식할 수 있었으나 점차 다양한 글씨체를 인식할 수 있게 되었고, 더 나아가 손 글씨까지 인식할 수 있는 수준으로 발전하였다(Eikvil, 1993; Deselaers et al., 2012). 이러한 OCR은 종이로 된 사무 문서를 컴퓨터에 입력하거나 스캔한 문서에서 텍스트를 추출할 때 등 여러 방면에서 유용하게 활용되고 있다.

그러나 이러한 OCR 기술은 현재 두 가지의 한계점을 지니고 있다. 첫 번째로는 외관상 비슷한 글자를 잘 구별하지 못한다는 점이다. 예를 들어, 숫자 0과 알파벳 O의 경우 두 개가 서로 유사하기 때문에 잘못 인식할 가능성이 크다. 두 번째로는 배경이 매우 어둡거나, 다른 글씨 혹은 그림 위에 써진 글씨들은 잘 인식하지 못한다는 점이다(Patel et al., 2012). 즉 기존 OCR 기술은 정형적인 문서 이미지를 인식할 때 좋은 성능을 보이지만, 다른 환경의 이미지 내 텍스트들을 인식하는 데에는 배경, 질감, 서식, 조명 등의 환경 영향 때문에 성능이 좋지 못할 수 있다(Jung et al., 2015). 이를 극복하기 위해 다양한 특징들을 수동으로 추출(Hand-crafted)하고 각각의 이미지 환경에 특화된 특징들을 선택하는 시도가 있었으나, 이미지 별로 문자 인식 성능의 편차가 크고 복잡한 문자의 특징을 잘 추출하지 못한다는 한계를 보였다(Jung et al., 2015; Yao et al., 2014).

이와 같은 기존의 OCR 기술들은 이미지 내 텍스트 전처리와 분류 알고리즘(Segmentation Algorithm)이 성능을 결정하였다(Patel et al., 2012). 하지만 인공지능이 딥러닝을 필두로 이미지 인식에서 두각을 드러내기 시작하자 이미지 내의 텍스트를 인식하기 위해 딥러닝을 사용하는 연구가 많아지고 있는데, 여러 논문들에서 인공신경망을 기반으로 OCR 프로그램을 제작하였을 때 기존보다 좋은 성능을 보인다는 결과가 보이고 있다(Singh, 2013). 본 연구 역시 카탈로그 데이터들이 일반 문서와는 다르게 다양한 서체와 각종 배경 이미지들로 꾸며져 있기에 기존 OCR 기술들의 한계를 넘고자 딥러닝을 활용해 키워드 인식 모형을 만들었다.

본 연구와 유사한 연구들로는 Gupta 등(2016)

이 이미지 내의 텍스트를 인식하기 위해 딥러닝 기법 기반 사물인식 모형 중 하나인 YOLO(You Only Look Once)를 개조하고, 이를 훈련시키기 위해 인공적으로 데이터를 생성한 연구 등을 들 수 있다. 그러나 기존 연구들이 인식기의 성능을 높이기 위해 모형의 구조를 바꾸는 것에 중점을 둔 것에 비해, 본 연구에서는 어떤 특징을 지닌 훈련데이터가 인식 모형을 효율적으로 학습시킬 수 있는지를 알아보고, 이를 통해 인식 모형의 성능을 높이고자 한다는 점에서 차이가 있다고 볼 수 있다.

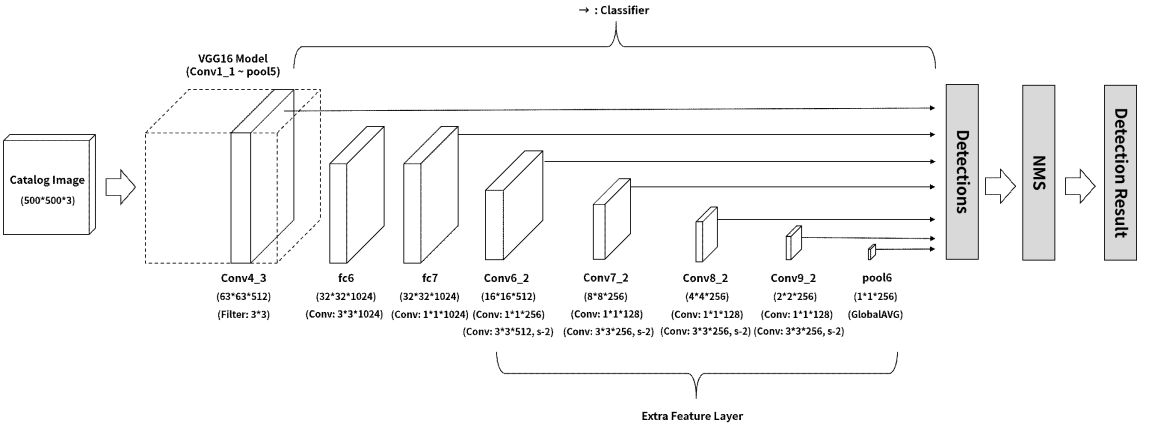
3. 키워드 텍스트 추출 방안 및 훈련 데이터 생성기 개발

3.1 딥러닝을 활용한 키워드 인식기의 개발

본 연구에서는 기존 OCR 프로그램 및 서비스의 한계로 인해 현재 이미지 인식 분야에서 좋은 성능을 보이고 있는 딥러닝 기법을 활용해 새로운 키워드 인식기 개발을 시도하였다. 새로운 검색 서비스 목표의 특성상 미리 지정된 키워드만을 인식해 텍스트로 변환하면 되는데, 이에 가장 적합한 딥러닝 모형은 이미지 내 사물 인식 모형이라 판단하였다. 사물 인식 모형은 이미지 내 사물의 특성을 학습하여 사물이 존재하는지 여부 뿐만이 아니라 어느 위치에 존재하는지도 인식이 가능하기에, 훈련 대상을 사물에서 키워드로 대체하면 충분히 한 카탈로그 이미지 안의 여러 키워드들을 동시에 인식할 수 있다고 판단하였다. 또한 기존 연구들에서도 사물 인식 모형을 활용하여 텍스트들을 인식한 경우가 있었다(Gupta et al., 2016). 본 연구는 기연구된 여러 가지 딥러닝 기법 기반 사물 인식 모형 중, 속도나 인식 성능 면에서 좋은 성능을 보인 SSD 모형을 기반으로 하여 키워드 인식기의 프로토타입을 개발하였다(Liu et al., 2016).

키워드 인식기의 프로토타입은 기존 SSD와 같이 크게 이미지를 입력 받는 입력 부분, 기존의 이미지 인식 모형인 VGG16 모형의 특징 추출 부분에 합성곱 층을 확장하여 만든 특징 추출 부분들, 특징 추출 부분에서 각 층 별로 추출된 특징들을 활용해 키워드들의 위치, 크기 및 종류를 직접 예측하는 검출 부분, 검출 부분에서 인식한 키워드들 중 중복하여 인식한 키워드들을 통합해주는 NMS(Non Maximum Suppression) 부분, 그리고 최종 인식 결과를 출력하는 출력 부분으로 이루어져 있다. 이를 도식적으로 표현하면 <Figure 1>과 같다.

본 키워드 인식기의 프로토타입에 대한 자세한 설명은 다음과 같다. 우선 입력 부분의 경우 일반적인 SSD 모형이 인식하는 사물들에 비해 본 키워드 인식기가 인지할 키워드들이 대부분 상대적으로 크기가 작다는 것을 고려하여, 상품 카탈로그 이미지를 입력할 때 500*500 픽셀로 변환한 이미지를 입력 받도록 설계하였다. 특징 추출 부분에서는 입력 받은 이미지들이 다수의 합성곱 층과 최대 풀링(Max Pooling) 층을 통과하면서 각 키워드들의 특징들을 차례대로 학습하게 구성되어 있다. 즉, 특징 추출 부분의 제일 앞 단에 있는 ‘Conv1_1’ 층에서는 키워드가 가지고 있는 직선이나 직각, 곡선과 같은 미시적인 특징들을 학습하며, 그보다 뒤에 있는 ‘Conv4_3’ 층에서는 앞 층에서 학습한 특징들을 기반으로 일반적인 키워드들의 전반적인 특징들을 학습하게 된다. 그리고 그보다 뒤에 있는 ‘Conv6_2’, ‘Conv7_2’, ‘Conv8_2’, ‘Conv9_2’ 등의 층에서는



〈Figure 1〉 Schematic diagram of the Keyword Recognition model based on SSD500 Model

이전 층들보다 크기가 큰 텍스트들의 특징들을 학습하도록 구성하였다. 그리고 마지막의 'pool6' 층의 경우 이미지 전체를 인지하도록 설계된 층으로, 합성곱 층 대신 특징 지도(Feature Map)들의 값들의 평균을 취하는 'Global Average Pooling' 층으로 구성되어 있다. 기존 SSD모형에서는 이 특징 추출 부분 중 VGG16 모형 부분에 해당하는 검출 부분을 구성할 때 기학습된 가중치를 불러온 뒤 훈련이 되지 않도록 설정하였지만, 본 연구에서는 사물과 글자의 특징이 다른 만큼 훈련이 이루어지도록 변경함과 동시에 학습률을 약 3배 정도 높은 0.001로 설정하였다.

다음으로 검출 부분에서는 앞서 설명한 'Conv4_3', 'fc7', 'Conv6_2', 'Conv7_2', 'Conv8_2', 'Conv9_2', 'pool6' 등 총 7개의 층에서 구한 특징 지도들의 특징들을 학습하는 합성곱 층과 미리 산정한 다수의 Default Box들의 정보들로 구성되어 있다. 이 부분을 통해 여러 위치에 존재하는 Default Box들이 어떠한 키워드인지, 얼마나 큰 텍스트인지 등을 산출하여 이미지 상에 키워드가 있을 영역들의 후보를 구하게 된다. 본 연구에서는 50% 이상의 확률을 가지는 영역들만을 검출하도록 설정하였다. 또한 사물 대비 텍스트들의 크기가 보통 작다는 것을 고려하여 Default Box들의 크기들을 기존 연구들보다 작게 설정하였다. 검출 영역에서 나온 결과물은 보통 한 키워드 주변에 중복되어 많은 Default Box들이 나타나게 되는데, NMS 부분에서는 이 중복되어 겹치는 박스들을 확률이 가장 높은 하나의 박스로 통합해 주는 역할을 담당하고 있다. 본 연구에서는 기연구된 모형과 같이 45% 이상 겹치는 부분이 존재하면 이를 통합하도록 하였다. 마지막으로 출력 부분에서는 NMS 부분에 의해 정리된 값들을 받아, 키워드의 종류와 확률, 상대좌표들로 구성되어 있는 결과값들을 키워드와 실제 이미지 상의 박스 좌표 값들을 산정하는 역할을 담당한다.

본 연구에서 사용된 키워드 인식기의 프로토타입에서는 널리 알려져 있는 SSD 모형과 유사한 설정들을 활용하였으나, 특징 추출 부분의 VGG16 모형을 다른 모형으로 변경하거나 검출 부분의 Default Box 및 필터 사이즈 등을 텍스트

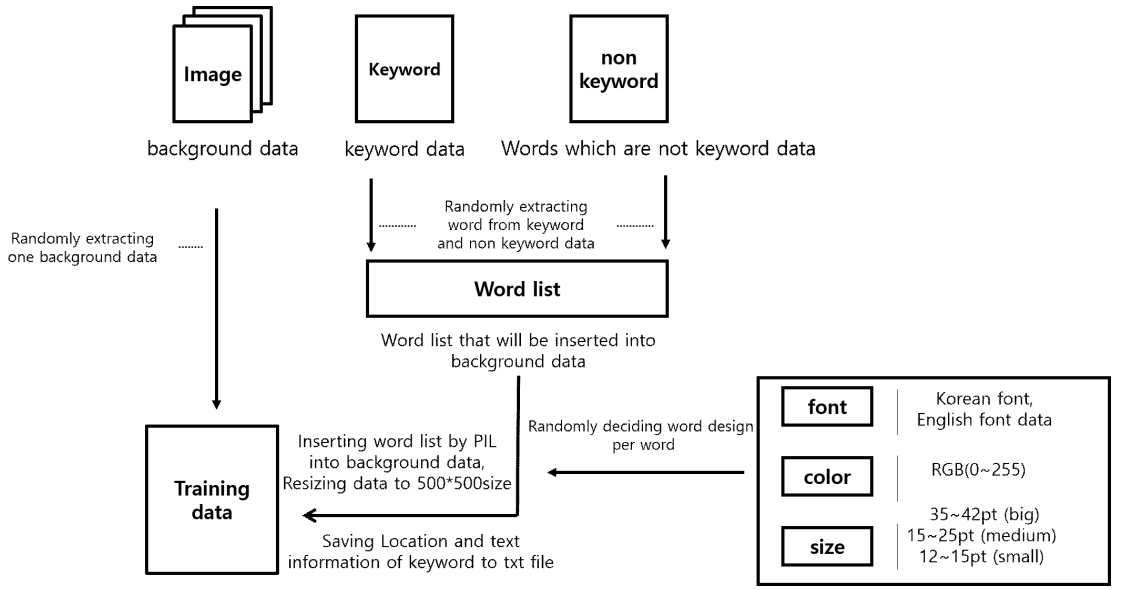
의 특징에 맞게 세부 조절함으로써 키워드 인식 성능을 향상시킬 수 있을 것이다.

3.2 훈련용 데이터 자동 생성기 개발

키워드 인식기 모형의 훈련을 위해서는 다수의 학습데이터를 확보해야한다. 하지만 키워드 인식기 모형에 사용되는 학습데이터는 인간이 일일이 데이터에 정답 라벨링을 해야 하기에 다수의 학습 데이터를 확보하는데 많은 시간과 비용이 소모된다. 또한 특정 라벨 값을 가진 학습 데이터만을 필요로 할 경우, 그러한 조건을 충족시키는 데이터를 수집하는 것도 어렵다.

본 연구에서는 학습데이터 정답 라벨링과 데이터 수집의 어려움을 해소하기 위한 방안으로 프로그램을 통해 학습 데이터를 생성하는 방법(훈련용 데이터 자동 생성기 개발)을 고안했다. 대다수 상품 카탈로그 이미지는 컴퓨터 환경 하에서 제작되었다는 점에서 학습데이터를 컴퓨터 프로그램으로 생성하여도 실제 카탈로그 이미지 데이터와 생산과정에서 유의적인 차이가 없다. 학습데이터를 프로그램을 통해 생성하면 원하는 조건의 데이터를 원하는 양 만큼 별도의 정답 라벨링 작업 없이 생산할 수 있다. 훈련용 데이터 자동 생성기는 Python 3.5 버전으로 개발되었으며 주요 사용 라이브러리는 'PIL(Python Image Library)'을 기반으로 한 'Pillow'와 내장 모듈인 'Random' 등이 있다. 본 생성기를 사용하여 데이터를 생성하는 과정은 아래 <Figure 2>와 같으며, 생성된 훈련데이터의 예는 <Figure 3>과 같다.

훈련용 데이터 자동 생성기를 개발한 결과, 원하는 특징을 가지는 데이터들을 빠른 시간 내에 원하는 양 만큼 쉽게 생산할 수 있게 되었다. 그러나 어떠한 특징의 데이터를 얼마나 생성해야 키워드 인식기 모형의 성능 향상에 도움이 되는지 명확하지 않아, 실험을 통해 이를 확인하였다.



〈Figure 2〉 Schematic diagram of progress of generating train data

〈Figure 3〉 Examples of generated train data

4. 실험 설계 및 결과

4.1 실험 목적

본 연구는 3장에서 제시한 키워드 인식기 및 훈련용 데이터 자동 생성기의 효과성을 실험하기 위해, 두 번의 실험을 설계하였다. 첫 번째로는 훈련 데이터 자동 생성기로 학습된 키워드 인식기의 성능을 기존 OCR 프로그램들의 성능과 비교하는 실험이다. 이를 통해 기존 OCR 프로그램들을 활용하여 상품 카탈로그 이미지들에서 키워드를 인식할 때의 한계와 함께, 본 키워드 인식기의 프로토타입이 가지는 강점에 대해 알아본다.

두 번째는 훈련용 데이터 자동 생성기로 어떠한 특징의 데이터를 생성해야 키워드 인식기 성능 향상에 효과적인지 알아보고자, 서로 다른 특징을 가진 데이터들로 훈련된 키워드 인식기의 성능을 비교하는 실험이다. 본래 지도학습에서는 기본적으로 훈련용 데이터와 시험용 데이터가 같은 종류인 것이 일반적이나, 본 연구에서 사용되는 훈련용 데이터는 상품 카탈로그 데이터를 이용하는 것이 아니라 임의로 제작하기 때문에 어떻게 제작하는지에 따라 성능 차이가 발생할 수 있다. 그렇기에 본 실험에서는 어떻게 훈련용 데이터를 제작하면 적은 양의 데이터로도 높은 성능을 발휘할 수 있는지 여러 방법들을 비교하면서 알아보는 것을 목표로 한다.

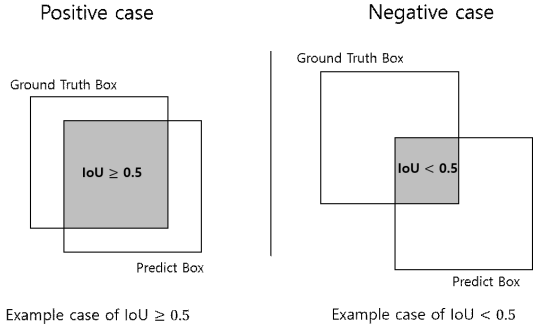
4.2 훈련용 데이터 자동 생성기를 통해 학습한 키워드 인식기의 성능 평가

4.2.1 키워드 인식기의 성능 평가 실험 절차

키워드 인식기의 성능을 평가하기 위해 다음과 같은 훈련 및 평가 과정을 거쳤다. 먼저 주방용품 카탈로그 이미지에서 자주 등장하는 10개의 단어를 키워드 인식기가 학습할 키워드로 선정하고, 훈련용 데이터 자동 생성기로 학습할 키워드들이 포함된 학습데이터 20,000장을 생성하였다. 다음으로 훈련할 학습데이터의 개수를 달리하며 키워드 인식기를 20번의 훈련횟수(epoch)로 학습시키고, 학습데이터 개수에 따른 키워드 인식기의 인식 성능을 평가하였다.

키워드 인식 성공의 기준은 키워드의 위치 정

보를 성공적으로 예측하는 동시에 올바르게 키워드를 분리하는 경우로 정하였다. 위치 정보를 예측한 결과는 정답으로 기록된 키워드의 영역(Ground Truth Box)과 인식기가 예측한 키워드의 영역(Predicted Bounding Box)의 합집합과 교집합의 비율인 IoU(Intersection over Union)가 0.5이상일 때 성공적으로 예측하였다고 인정하였다(<Figure 4> 참조). 이는 PASCAL VOC challenge(Everingham et al., 2010) 등 기존 사물 인식 분야에서 사용하는 평가 기준을 따른 것으로, 키워드 인식 성공 기준을 만족하는 키워드들은 키워드 인식기의 성능을 평가할 때 키워드 인식기가 맞춘 정답으로 계산하였다.



〈Figure 4〉 Explanation of recognition success standard with IoU

키워드 인식기의 성능 평가 기준에는 정밀도(Precision)와 재현율(Recall)의 조화 평균인 F-Score 지표를 사용하였다(<Equation 1> 참조). 정밀도는 키워드 인식기가 예측한 키워드의 개수 중 정답을 맞춘 비율을 의미하고, 재현율은 전체 키워드의 개수 중 정답을 맞춘 비율을 의미한다. 성능 평가를 위한 테스트 데이터는 사람이 직접 정답 라벨링 작업을 한 500*500 픽셀 크기의 카탈로그 이미지 150장(이하 시험이미지)을 이용하였다.

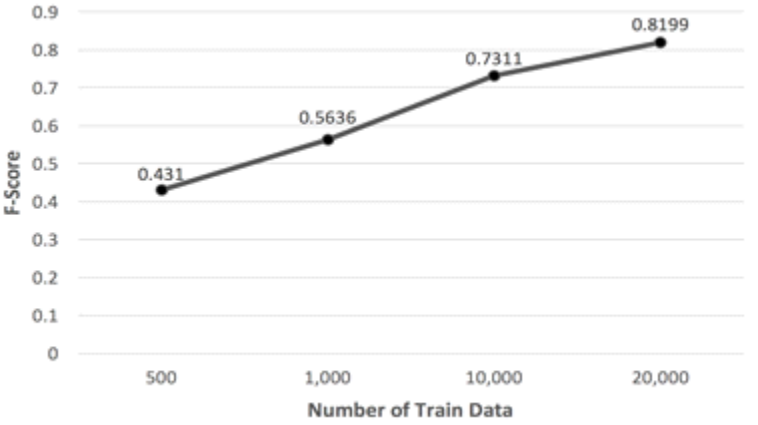
$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad \text{<Equation 1>}$$

키워드 인식기와 기존 OCR프로그램의 성능과 비교하기 위해서 현재 출시되어 있는 OCR프로그램 중 일부를 선정하여 성능을 평가했다. 기존 OCR프로그램으로는 최근 인공지능 영역에서 우수한 성과를 보이고 있는 Google사의 무료 OCR서비스인 Google Docs의 OCR 서비스와, 세계 수준의 OCR기술을 가지고 있다고 평가받는 ABBYY사의 엔진을 사용한 ABBYY Fine Reader 14를 선정하였다(Kim, 2016). 테스트 데이터는 키워드 인식기의 성능을 평가할 때와 동일하게 시험이미지 150장을 이용했다. 다만 기존 OCR 프로그램들의 경우 인식 결과물에서 위치 정보를 활용할 수 없어 평가 기준은 부득이 재현율만을 사용하였고, 인식성공 기준은 목표로 하는 키워드가 추출되는 경우로 한정하였다.

4.2.2 키워드 인식기의 성능 평가 실험 결과

키워드 인식기의 성능 평가를 위한 실험 결과, 학습데이터의 개수가 500개, 1,000개일 때 각각 0.4310, 0.5636의 F-score, 10,000개일 때 0.7311의 F-score, 20,000개일 때 0.8199의 F-score를 기록하였다(<Figure 5> 참조). 이를 통해 학습데이터가 많을수록 키워드 인식기의 성능은 향상되지만, 성능 향상 폭은 점차 작아짐을 볼 수 있었다. 20,000장의 훈련용 데이터로 훈련된 인식기의 인식 결과 예시는 아래 <Figure 6>과 같다.

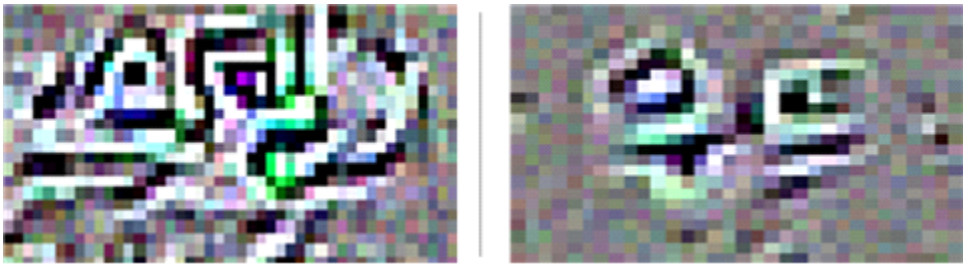
이후 실제로 키워드 인식기 모형이 글자들의 특징을 학습하였는지 확인해보고자, 필터의 시



〈Figure 5〉 Performance of the Keyword Recognition model depending on number of train data

〈Figure 6〉 Example of recognizing keywords with Keyword Recognition model

각화를 시도하였다. 그 결과 키워드 인식기가 학습한 키워드 글자 모양들의 특징들을 발견할 수 있었다. 아래 <Figure 7>은 20,000개로 훈련시킨 키워드 인식기가 보여준 여러 특징들 중 '스틸'과 '우드'라는 키워드를 시각화 한 결과이다.



〈Figure 7〉 Feature maps of keywords extracted by the Keyword Recognition model

다음으로 기존 OCR 프로그램들의 성능 평가 결과, Google Docs의 OCR프로그램은 45.78%, ABBYY의 Fine Reader 14는 51.92%의 재현율을 보였다. 키워드 인식기와의 성능 비교를 위해 평가 기준을 재현율로 설정할 때, 키워드 인식기의 재현율은 20,000개의 데이터로 훈련시켰을 경우 73.91%로, 위치정보를 고려하여 보다 엄격한 기준임에도 불구하고 기존 OCR프로그램의 최대 재현율인 51.92%보다 우수한 성능을 보였다 (<Table 1> 참조). 이러한 성능 비교를 통해 본 연구의 키워드 인식기가 기존 OCR프로그램보다 카탈로그 이미지에서 특정 키워드 검출에서 보다 나은 성능을 보임을 확인하였다. 다만 본 실험의 경우 이미지 상의 특정한 키워드만을 찾아낸다는 문제에 한정한 경우로, 모든 글자에 대해 범용적으로 동작하는 기존 OCR 프로그램들에 대해 항상 우수한 성능을 낸다고 보기 어려운 면이 있다. 그러나 본 인식기의 성능과 비교할 기존 연구들의 부재로 부득이 본 실험과 같이 진행하였으며, 차후 본 연구를 기반으로 키워드의 수, 종류 등을 달리하여 인식 성능을 향상시키는 추가적인 연구가 필요할 것이다.

4.3 생성된 데이터 특징에 따른 키워드 인식기 성능 비교

4.3.1 데이터 특징에 따른 성능 비교 실험 절차

딥러닝의 경우 훈련시 조절할 수 있는 변수의 수가 매우 많아 모든 변수에 대하여 실험을 진행할 수 없었다. 이에 부득이 본 실험을 위해 훈련용 데이터 자동 생성기와 키워드 인식기 모형을 다소 수정하였다. 훈련용 데이터 자동 생성기의 경우, 데이터 생성 시 무작위로 결정되는 부분의 범위를 통제하거나 생성 프로그램의 입력 값들

〈Table 1〉 Performances of existing OCR programs and keyword recognition model

	Number of Total Keywords	Number of Extracted Keywords	Recall Ratio (%)
Google Docs OCR	391	179	45.78%
ABBYY Fine Reader 14	391	203	51.92%
Keyword Recognition model	391	289	73.91%

을 바꿔 실험 조건에 맞게 데이터를 생성하도록 변경하였다. 그리고 키워드 인식기 모형에서는 기존 SSD 모형에서 추가적인 데이터 확보를 위해 사용되던 임의 밝기 조절이나 이미지 기울이기, 확대 등과 같은 데이터 확장(Data Augmentation) 방법은 실험의 통제를 위해 사용하지 않았다.

1,000개의 데이터를 기준으로 각 데이터에 투입되는 정답 라벨을 증가시켰을 경우(A case), Bounding Box의 가로 세로 크기를 수정하여 같은 위치에 정답 라벨을 늘릴 경우(B case), 정답 라벨이 아닌 단어를 없앨 경우(C case), 단어 간 공백을 다르게 할 경우(D case), 배경이미지를 다르게 할 경우(E case) 성능의 차이가 존재하는지 실험했으며, 5가지 경우에 따른 성능차이를 실험하였다. 5가지 경우에 대한 상세 정보는 다음 <Table 2>와 같다.

5개의 케이스를 조합한 72개의 데이터 셋을 5번 만든 뒤, 각 케이스별 성능차가 존재하는지를 실험하였다. 실험과정에서 각 데이터 셋은 훈련용 데이터 자동 생성기로 <Table 2>의 제약 조건에 맞춰 각각 1,000개씩 생성하였으며, 이 중 랜덤하게 800개는 훈련용 데이터로, 200개는 검증용 데이터로 선정하였다. 테스트용 데이터로는 직접 만든 온라인 주방용품 사이트의 실제 카탈로그 이미지 150장을 선정하였다. 다만 본 실험을 진행하며 딥러닝의 특성상 파라미터가 많아 훈련 시간이 오래 걸림에도 불구하고 360개나 되는 모형들을 훈련시켜야 하는 만큼, 한 모형

〈Table 2〉 Explanation of 5 cases in experiment

Case		Explanation
A	0	Randomly deciding the number of keyword label per data from normal distribution (mean : 5, var : 3)
	1	Randomly deciding the number of keyword label per data from normal distribution (mean : 10, var : 3)
B	0	Placing one keyword label on keyword location
	1	Placing a keyword label five times of which bounding boxes are randomly adjusted ± 3pixel on keyword location
C	0	Inserting 40~60 of words which are not keyword
	1	Not inserting words which are not keyword
D	0	Space between words is 3 pixel
	1	Space between words is 40 pixel
	2	Space between words is randomly decided from normal distribution (mean : 5, var : 2)
E	0	Using background of which text area is solid color image, other area is catalog image
	1	Using solid color background image
	2	Using catalog background image

당 수 만 번의 훈련 차수(epoch)를 진행하기에는 여러가지 어려움이 존재하였다. 또한 본 실험은 최고의 모형을 가리는 것이 아니며, 같은 조건 하에서 어떠한 조건의 데이터들이 더 좋은 성능을 발휘하는지 점검하는 것이기에 훈련 횟수를 다소 낮게 설정하기로 결정하였다. 이에 다수의 실험 결과, 본 키워드 인식기의 인식 성능은 평균적으로 20회 이내의 훈련 차수에서 거의 최고 수준에 다다른다는 것을 확인하고 이를 기준으로 훈련을 진행하였다. 이러한 방식으로 총 360개의 데이터 셋을 훈련시킨 후 성능을 평가했으며 성능 평가의 지표는 정밀도(Precision)와 재현율(Recall)의 조화 평균인 F-Score 지표를 사용했다. 각 케이스별 F-Score를 토대로 대응표본 t-test를 하여 데이터 특성에 따른 F-Score에 유의적인 차이가 있는지 검정하였다. 표본 집단 간의 대응을 위해 데이터 생성 시 무작위로 결정되는 부분은 Python 내 random 모듈의 시드값(Seed Number)을 이용하여 통제했다.

4.3.2 데이터 특징에 따른 성능 비교 실험 결과

A Case에 대한 실험 결과는 아래 <Table 3>과 같다. 분석 결과, 단측 유의확률이 0.0000으로 유의수준 0.01에서 두 집단간 유의한 차이가 있었다. 이는 이미지 당 투입되는 정답 라벨의 수를 평균 5개에서 평균 10개로 증가시키면 인식 성능이 높아진다는 것을 의미하며, 같은 양의 데이터를 활용할 때 더 많은 키워드들의 특징을 학습했다고 볼 수 있다.

B Case에 대한 실험 결과는 <Table 4>와 같다. 분석 결과, 단측 유의확률이 0.0000으로 유의수준 0.01에서 두 집단간 유의한 차이가 있었다. 이를 통해 키워드의 위치에 Bounding Box의 크기를 수정하여 정답 라벨을 1개에서 5개로 늘리는 것이 성능 향상에 도움이 된다는 것을 확인할 수 있다. 이에 대해서는 일반적으로 크기가 작고 다양한 형태를 지니는 키워드들에 대해 다양한 정답 모양을 줌으로써, 한 가지 형태의 정답만을 줄 때 보다 여러가지 형태의 Default Box들이 키워드들의 각기 다른 특징들을 학습할 수 있게 도움을 주었다고 본다.

C Case에 대한 실험 결과는 <Table 5>와 같다. 분석 결과, 단측 유의확률이 0.0000으로 유의수준 0.01에서 두 집단간 유의한 차이가 있었다. 이는 정답 라벨이 아닌 단어를 투입하는 것이 그렇지 않은 경우보다 성능 향상에 도움이 된다는 것을 의미한다. 이는 정답이 아닌 단어

〈Table 3〉 Paired T-Test Result for A Case

	N	Mean		Std. Deviation	
A0	180	0.0578		0.1142	
A1	180	0.1730		0.1663	
Paired Differences			t	df	Sig. (1-tailed)
Mean	Std. Deviation	Std. Error Mean			
-0.1151	0.1538	0.0114	-10.0406	179	0.0000***

*** p < 0.01 / ** p < 0.05 / * p < 0.1

〈Table 4〉 Paired T-Test Result for B Case

	N	Mean		Std. Deviation	
B0	180	0.0871		0.1271	
B1	180	0.1437		0.1720	
Paired Differences			t	df	Sig. (1-tailed)
Mean	Std. Deviation	Std. Error Mean			
-0.0566	0.1197	0.0089	-6.3389	179	0.0000***

*** p < 0.01 / ** p < 0.05 / * p < 0.1

〈Table 5〉 Paired T-Test Result for C case

	N	Mean		Std. Deviation	
C0	180	0.1799		0.1915	
C1	180	0.0509		0.0480	
Paired Differences			t	df	Sig. (1-tailed)
Mean	Std. Deviation	Std. Error Mean			
0.1290	0.1628	0.0121	10.6250	179	0.0000***

*** p < 0.01 / ** p < 0.05 / * p < 0.1

들을 키워드 인식기가 같이 학습하면서 정답인 키워드들에 대해 보다 강건한 특징들을 학습했다고 볼 수 있다.

D Case에 대한 실험 결과는 <Table 6>과 같다. 3건의 대응 중 단측 유의확률이 0.01 이내인 대응은 D0-D1 대응과 D1-D2 대응이었다. 이에 유의수준 0.01에서 단어 간 공백이 3픽셀인 경우와 단어 간 공백이 평균 5, 분산 2픽셀인 정규분포에서 무작위로 정해질 경우가 단어 간 공백이 40픽셀일 경우보다 성능 향상에 더 도움이 된다고 볼 수 있다. 다만 D0-D2 대응은 유의 확률이 0.1220으로 유의수준 0.1에서 인식 성능에 유의한 차이가 있다고 볼 수 없었다. 이는 Default Box를 사용하여 목표 사물 주변부의 영향을 받는 SSD 모형의 특성상, 정답인 키워드의 주변부까지 같이 학습하는 현상을 방지하였다고 볼 수 있다.

E Case에 대한 실험 결과는 <Table 7>과 같다. E0-E1 대응과 E0-E2 대응은 각각 단측 유의확률이 각각 0.0074, 0.0024로, 유의수준 0.01에서 글자 영역은 단색이고 나머지 영역은 카탈로그 이미지인 배경 데이터가 단색 배경 또는 카탈로그 이미지 배경인 데이터보다 성능 향상에 도움이 된다고 볼 수 있다. 다만 E1-E2 대응은 단측 유

〈Table 6〉 Paired T-Test Result for D case

	N	Mean	Std. Deviation
D0	120	0.1556	0.1680
D1	120	0.0240	0.0222
D2	120	0.1666	0.1726

Pair	Paired Differences			t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean			
D0 – D1	0.1316	0.1628	0.0142	9.2560	119	0.0000***
D0 – D2	-0.0110	0.1023	0.0093	-1.1740	119	0.1220
D1 – D2	-0.1426	0.1605	0.0147	-9.7323	119	0.0000***

*** p < 0.01 / ** p < 0.05 / * p < 0.1

〈Table 7〉 Paired T-Test Result for E case

	N	Mean	Std. Deviation
E0	120	0.1331	0.1670
E1	120	0.1102	0.1503
E2	120	0.1029	0.1423

Pair	Paired Differences			t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean			
E0 – E1	0.0229	0.1012	0.0092	2.4760	119	0.0074***
E0 – E2	0.0301	0.1143	0.0104	2.8873	119	0.0024***
E1 – E2	0.0073	0.1191	0.0109	0.6679	119	0.2527

*** p < 0.01 / ** p < 0.05 / * p < 0.1

의확률이 0.2527로 유의수준 0.1에서 인식 성능에 유의한 차이가 있다고 볼 수 없었다. 이는 본 실험의 시험 데이터인 상품 카탈로그 이미지의 특성상 글자 이외에 각종 사물들의 이미지들이 많이 들어가게 되는데, 이러한 사물 이미지들은 키워드의 일부가 아님을 훈련 단계에서 학습시키는 것이 중요하다고 볼 수 있다.

5. 결론

본 연구에서는 전자상거래에서의 검색서비스 향상을 위해 상품 카탈로그 이미지 내 키워드를 인식하는 딥러닝 기반 키워드 인식기 모형을 제시했다. 그와 동시에 키워드 인식기 모형의 학습 데이터 수집 및 정답 라벨링 작업의 문제점을 해결하고자 훈련용 데이터 자동 생성기를 고안하였다. 훈련용 데이터 자동 생성기의 데이터를 학습한 키워드 인식기는 기존 OCR 프로그램보다 높은 키워드 인식 성능을 보여 상품 카탈로그 이미지 내 텍스트 인식에 본 연구의 딥러닝 기반 인식 모형이 효과적임을 보였다. 또한 본 연구는 훈련용 데이터 자동 생성기가 어떠한 특징의 데이터를 생성해야 키워드 인식기의 훈련에 효과적인지를 알아보는 실험을 수행하였다. 서로 다른 특징의 데이터들을 각각 1,000개씩 생성하여 키워드 인식 모형을 훈련시켰고, 각각의 인식 성능을 측정하였다. 이를 대응 표본 t-test를 통해 분석한 결과, 이미지 당 들어가는 정답 라벨의 수를 증가시키거나, Bounding box의 크기만을 소폭 조정하여 같은 위치에 정답 라벨의 수를 늘리는 것이 좀 더 키워드 인식기의 학습에 효과적임을 밝혔다. 그리고 학습시킬 키워드뿐만 아니라 학습 대상이 되지 않는 단어를 학습데이터에 넣는 것, 단어 간 공백을 넓게 하지 않는 것, 글자 영역은 단색, 그 외 영역은 카탈로그 이미지의 배경과 동일한 배경 이미지를 쓰는 것이 키워드 인식기 모형 훈련에 더욱 효과적임을 보였다. 이러한 결과들을 통해 본 연구가 키워드 인식기를 통한 전자상거래의 검색 영역의 개선뿐만 아니라, 차후 기계를 사용한 한국어 텍스트 인식 시 필요한 데이터 확보 등에서 많은 연구자들에게 도움을 줄 수 있을 것으로 예상한다.

본 연구의 한계로는 딥러닝의 훈련에 영향을 끼치는 모든 변수에 대해 실험 장비 및 시간의 제약으로 실험할 수 없어, 이보다 더 좋은 결과를 보여주는 조합이 있을 가능성이 있다는 것이다. 다음으로 본 연구에서는 10개라는 다소 작은 수의 키워드만을 사용하여 효과적인 학습데이터 생성방안을 제시하였기에, 키워드의 개수를 크게 증가시킬 경우 이를 설명하지 못한다는 한계를 지니고 있다. 그렇기에 후속 연구로는 키워드 수를 확장하여 키워드 인식기 및 훈련용 데이터 자동 생성기의 확장성(scalability)에 대한 검토를 진행할 예정이다. 이에 대해서는 현재 키워드 인식 모형과 훈련용 데이터 생성기를 개선한 결과, 키워드의 개수를 50개로 늘렸을 경우에도 우수한 인식 성능을 얻을 수 있음을 확인하였으며 이를 바탕으로 차후 200개 이상의 키워드에도 대응할 수 있도록 차후 연구를 계속 진행할 예정이다.

참고문헌(References)

Cao, G., X. Xie, W. Yang, Q. Liao, G. Shi, and J. Wu, “Feature-Fused SSD: Fast Detection for Small Objects,” *arXiv preprint*, (2017).

Cho, S. Y., J. E. Choi, K. H. Lee, and H. W. Kim, “An online review mining approach to a recommendation system,” *Information Systems Review*, Vol.17, No.3(2015), 95~111.

Choi, H. Y., and Y. H. Min, “Introduction to deep learning and major issues[written in Korean],” *Korea Information Processing Society Review*, Vol.22, No.1(2015), 1-15.

Choi, S. I., Y. J. Hyun, and N. G. Kim, “Improving performance of recommendation

systems using topic modeling," *Journal of Intelligence and Information Systems*, Vol.21, No.3(2015), 101~116.

Deselaers, T., T. Gass, G. Heigold, and H. Ney, "Latent log-linear models for handwritten digit classification," *IEEE transactions on pattern analysis and machine intelligence*, Vol.34, No.6(2012), 1105~1117.

Everingham, M., L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, Vol.88, No. 2(2010), 303-338.

Eikvil, L., "Optical character recognition," *Technical Report*, Norwegian Computing Center, 1993.

Fu, C. Y., W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional Single Shot Detector," *arXiv preprint*, (2017).

Girshick, R., "Fast r-cnn," *The IEEE International Conference on Computer Vision (ICCV)*, (2015), 1440-1448.

Girshick, R., J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2014), 580-587.

Gupta, A., A. Vedaldi, and A. Zisserman, "Synthetic data for text localisation in natural images," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 2315~2324.

Hong, M. D., J. W. Kim, and G. S. Jo, "A wordnet-based open market category search system for efficient goods registration," *Journal of the Korea society of computer and information*, Vol.17, No.9(2012), 17~27.

Hwang, C. G., M. N. Yi, and G. D. Jung, "Design of merchandise retrieval system based on ontology on EC," *Proceedings of the Korean Society for Internet Information*, Vol.6, No.1(2005), 213~216.

Jung, K. H., H. J. Kim, and Y. H. Lee, "Character recognition in general video using deep learning[written in Korean]," *Korea Information Processing Society Review*, Vol.22, No.1(2015), 42~54.

Kim, H. A., *Free 'ROSE document recognition', image to excel conversion function added*[written in Korean], EDAYIL, 2016. Available at http://www.edaily.co.kr/news/NewsRead.edy?newsid=01466166612883112 (Accessed 13 July, 2017)

Kim, H. J., "Dynamic hand gesture recognition using CNN model and FMM neural networks," *Journal of Intelligence and Information Systems*, Vol. 16, No. 2(2010), 95-108.

Kim, J. W., H. A. Pyo, J. W. Ha, C. K. Lee, and J. H. Lee, "Deep learning algorithms and applications," *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 33, No. 8(2015), 25-31.

Kim, K. J., B. G. Kim, "Product recommender system for online shopping malls using data mining techniques," *Journal of Intelligence and Information Systems*, Vol.11, No.1(2005), 191~205.

Kim, K. S., "A hybrid collaborative filtering algorithm for personalized recommendations and its application to the internet electronic commerce," *The Journal of Internet Electronic Commerce Research*, Vol.8,

No.4(2008), 1~20.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolution neural networks," *Advances in neural information processing systems*, Vol.25(2013), 1097-1105.

LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, Vol. 1, No. 4(1989), 541-551.

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *arXiv preprint*, (2016).

Ma, J., I. H. Jeon, and Y. K. Choi, "Design of an efficient keyword-based retrieval system using concept lattice," *Journal of Internet Computing and Services*, Vol.16, No.3(2015), 43~57.

Minsky, M., and S. Papert, *Perceptrons*. M.I.T. Press, Oxford, England, 1969.

Mo, Y. I., and C. G. Lee, "A study on increasing the efficiency of image search using image attribute in the area of content-based image retrieval," *Journal of the Korea society for simulation*, Vol.18, No.2(2009), 39~48.

Patel, C., A. Patel, and D. Patel, "Optical character recognition by open source OCR tool tesseract: A case study," *International Journal of Computer Applications*, Vol.55, No.10(2012), 50~56.

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779~788.

Ren, S., k. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, (2015), 91-99.

Rosenblatt, F., "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological review*, Vol.65, No.6(1958), 386~408.

Singh, S., "Optical character recognition techniques: a survey," *Journal of emerging Trends in Computing and information Sciences*, Vol.4, No.6(2013), 545~550.

Yang, G. M., *E-commerce industry to attract investment attraction 'hot'... The market gets bigger*.[written in Korean], NEWSIS, 2017. Available at http://www.newsis.com/view/?id=NISX20170426_0014856681 (Accessed 13 July, 2017).

Yang, J. G., S. I. Kwon, and Y. M. Yu, "A study on the current state of cross-border e-commerce and strategic activation plans for overseas direct sales," *E-Trade Review*, Vol.14, No.1(2016), 23~46.

Yao, C., X. Bai, and W. Liu, "A unified framework for multioriented text detection and recognition," *IEEE Transactions on Image Processing*, Vol.23, No.11(2014), 4737~4749.

Zhang, B. T., "Deep Hypernetwork Models," *Communications of the Korean Institute of Information Scientists and Engineers*, Vol.33, No.8(2015), 11-24.

Abstract

The way to make training data for deep learning model to recognize keywords in product catalog image at E-commerce

Kitae Kim* · Wonseok Oh** · Geunwon Lim** · Eunwoo Cha** · Minyoung Shin*** · Jongwoo Kim****

From the 21st century, various high-quality services have come up with the growth of the internet or 'Information and Communication Technologies'. Especially, the scale of E-commerce industry in which Amazon and E-bay are standing out is exploding in a large way. As E-commerce grows, Customers could get what they want to buy easily while comparing various products because more products have been registered at online shopping malls.

However, a problem has arisen with the growth of E-commerce. As too many products have been registered, it has become difficult for customers to search what they really need in the flood of products. When customers search for desired products with a generalized keyword, too many products have come out as a result. On the contrary, few products have been searched if customers type in details of products because concrete product-attributes have been registered rarely.

In this situation, recognizing texts in images automatically with a machine can be a solution. Because bulk of product details are written in catalogs as image format, most of product information are not searched with text inputs in the current text-based searching system. It means if information in images can be converted to text format, customers can search products with product-details, which make them shop more conveniently.

There are various existing OCR(Optical Character Recognition) programs which can recognize texts in images. But existing OCR programs are hard to be applied to catalog because they have problems in

* R&D Center, Mindgroup
** School of Business, Hanyang University
*** Department of Chinese Language & Literature, Hanyang University
**** Corresponding Author: Jongwoo Kim
School of Business, Hanyang University
222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea
Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail: kjw@hanyang.ac.kr

recognizing texts in certain circumstances, like texts are not big enough or fonts are not consistent. Therefore, this research suggests the way to recognize keywords in catalog with the Deep Learning algorithm which is state of the art in image-recognition area from 2010s. Single Shot Multibox Detector(SSD), which is a credited model for object-detection performance, can be used with structures re-designed to take into account the difference of text from object. But there is an issue that SSD model needs a lot of labeled-train data to be trained, because of the characteristic of deep learning algorithms, that it should be trained by supervised-learning. To collect data, we can try labelling location and classification information to texts in catalog manually. But if data are collected manually, many problems would come up. Some keywords would be missed because human can make mistakes while labelling train data. And it becomes too time-consuming to collect train data considering the scale of data needed or costly if a lot of workers are hired to shorten the time. Furthermore, if some specific keywords are needed to be trained, searching images that have the words would be difficult, as well.

To solve the data issue, this research developed a program which create train data automatically. This program can make images which have various keywords and pictures like catalog and save location-information of keywords at the same time. With this program, not only data can be collected efficiently, but also the performance of SSD model becomes better. The SSD model recorded 81.99% of recognition rate with 20,000 data created by the program.

Moreover, this research had an efficiency test of SSD model according to data differences to analyze what feature of data exert influence upon the performance of recognizing texts in images. As a result, it is figured out that the number of labeled keywords, the addition of overlapped keyword label, the existence of keywords that is not labeled, the spaces among keywords and the differences of background images are related to the performance of SSD model. This test can lead performance improvement of SSD model or other text-recognizing machine based on deep learning algorithm with high-quality data.

SSD model which is re-designed to recognize texts in images and the program developed for creating train data are expected to contribute to improvement of searching system in E-commerce. Suppliers can put less time to register keywords for products and customers can search products with product-details which is written on the catalog.

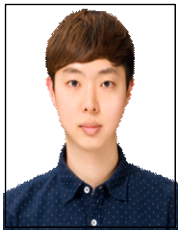
Key Words : Deep learning, train data generation, OCR, attribute-based search, Single Shot MultiBox Detector

Received : July 21, 2017 Revised : December 18, 2017 Accepted : January 14, 2018
Publication Type : Regular Paper(Fast-track) Corresponding Author : Jongwoo Kim

저 자 소 개

김 기 태

현재 ㈜마인드그룹의 개발본부에서 연구원으로 재직 중이다. 한양대학교에서 경영학사와 정보시스템학사를 취득하였으며, 한양대학교 일반대학원에서 경영정보시스템을 전공하여 경영학 석사학위를 취득하였다. 주요 연구 관심분야는 기계학습 및 딥러닝 기법의 비즈니스 활용, 데이터마이닝, 빅 데이터 등이다.



오 원 석

현재 한양대학교 경영대학 IBIS(Intelligent Business Information System) 연구실에서 연구원으로 재직 중이다. 한양대학교 경영학과 졸업 예정이며, 한양대학교 빅 인텔리전트 경영교육 사업단에서 주관하는 프로젝트 학기제 인턴 과정을 수료하였다. 주요 관심분야는 빅데이터, 디지털 마케팅, 딥러닝을 이용한 이미지 인식 등이다.

임 근 원

현재 한양대학교 경영대학 IBIS(Intelligent Business Information System) 연구실에서 연구원으로 재직 중이다. 한양대학교 경영학과 졸업 예정이며, 한양대학교 빅 인텔리전트 경영교육 사업단에서 주관하는 프로젝트 학기제 인턴 과정을 수료하였다. 주요 관심분야는 모바일 어플리케이션 개발, 웹 개발, 기계학습 및 딥러닝 등이다.

차 은 우

현재 한양대학교 경영학부 학생으로 재학 중이다. 한양대학교 빅 인텔리전트 경영교육 사업단에서 주관하는 프로젝트 학기제 인턴 과정에서 A.I.랩에 근무하였다. 주요 관심분야는 감성분석, 데이터 마이닝, 기계학습 및 딥러닝이다.

신 민 영

현재 한양대학교 인문과학대학 중어중문학과에 재학 중이다. 한양대학교 ㈜한양비즈랩에서 주관하는 인턴과정을 수료하였다(Business AI Lab). 주요 연구 관심분야는 E-commerce 추천 기술, 빅 데이터, 딥러닝 기술 등이다.

김 종 우

현재 한양대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교 수학과에서 학사를 마쳤으며, 한국과학기술원에서 경영과학으로 석사학위를, 산업경영학으로 박사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝 기법과 응용, 기계학습과 딥러닝, 오피니언 마이닝, 상품추천기술, 지능형 정보시스템, 집단지성, 사회 네트워크 분석, 클라우드 컴퓨팅 서비스 등이다.