

A Lightweight Deep Learning Model for Text Detection in Fashion Design Sketch Images for Digital Transformation

Ju-Seok Shin*, Hyun-Woo Kang*

*Professor, Dept. of Artificial Intelligence, Korea Polytechnics, Daegu, Korea

*Professor, Dept. of Artificial Intelligence, Korea Polytechnics, Daegu, Korea

[Abstract]

In this paper, we propose a lightweight deep learning architecture tailored for efficient text detection in fashion design sketch images. Given the increasing prominence of Digital Transformation in the fashion industry, there is a growing emphasis on harnessing digital tools for creating fashion design sketches. As digitization becomes more pervasive in the fashion design process, the initial stages of text detection and recognition take on pivotal roles. In this study, a lightweight network was designed by building upon existing text detection deep learning models, taking into consideration the unique characteristics of apparel design drawings. Additionally, a separately collected dataset of apparel design drawings was added to train the deep learning model. Experimental results underscore the superior performance of our proposed deep learning model, outperforming existing text detection models by approximately 20% when applied to fashion design sketch images. As a result, this paper is expected to contribute to the Digital Transformation in the field of clothing design by means of research on optimizing deep learning models and detecting specialized text information.

▶ **Key words:** Deep Learning, Lightweight Deep Learning Model, Fashion Design Sketch, Text Detection, Digital Transformation

[요 약]

본 논문에서는 의류 디자인 도면 이미지의 글자 검출을 위한 경량화된 딥러닝 네트워크를 제안하였다. 최근 의류 디자인 산업에서 Digital Transformation의 중요성이 대두되면서, 디지털 도구를 활용한 의류 디자인 도면 작성이 강조되고 있으며, 디지털화된 의류 디자인 도면의 활용 가능성을 고려할 때, 도면에서 글자 검출과 인식이 중요한 첫 단계로 간주된다. 이 연구에서는 기존의 글자 검출 딥러닝 모델을 기반으로 의류 도면 이미지의 특수성을 고려하여 경량화된 네트워크를 설계하였으며, 별도로 수집한 의류 도면 데이터 셋을 추가하여 딥러닝 모델을 학습시켰다. 실험 결과, 제안한 딥러닝 모델은 의류 도면 이미지에서 기존 글자 검출 모델보다 약 20% 높은 성능을 보였다. 따라서 이 논문은 딥러닝 모델의 최적화와 특수한 글자 정보 검출 등의 연구를 통해 의류 디자인 분야에서의 Digital Transformation에 기여할 것으로 기대한다.

▶ **주제어:** 딥러닝, 경량화 딥러닝 모델, 의류 디자인 도면, 글자 검출, 디지털 트랜스포메이션

- First Author: Ju-Seok Shin, Corresponding Author: Hyun-Woo Kang
- Ju-Seok Shin (jsshin7@kopo.ac.kr), Dept. of Artificial Intelligence, Korea Polytechnics
- Hyun-Woo Kang (hwkang@kopo.ac.kr), Dept. of Artificial Intelligence, Korea Polytechnics
- Received: 2023. 08. 25, Revised: 2023. 09. 11, Accepted: 2023. 10. 04.

I. Introduction

의류 디자인 도면은 의류 산업에서 중요한 역할을 담당하며, 디자이너들이 새로운 아이디어를 개발하고 제품화하는 데 있어 핵심적인 요소 중 하나이다. 더불어, 디지털 도구를 활용하는 기업들은 의류 디자인 및 생산 과정에서 높은 생산성과 효율성을 얻고 있는 추세이다[1-3]. 그러나 영세기업의 경우에는 아직까지 디지털 도구를 활용하지 않고 고전적인 방법으로 디자인 도면을 작성하고 있는 실정이다.

한편, 디지털 트랜스포메이션은 디지털 기술 및 디지털 도구를 기업의 비즈니스 프로세스에 통합하여 생산성과 경쟁력을 향상시키는 과정을 의미한다[4]. 특히, 디지털 트랜스포메이션을 적용한 기업들은 빠르게 변화하는 시장 환경에 더 적극적으로 대응하고, 더 나은 제품과 서비스를 개발하는 데에 성공적인 사례들을 많이 볼 수 있다[5]. 이러한 디지털 트랜스포메이션 기술의 도입은 의류 디자인 분야에서도 큰 영향을 미치고 있으며, 기업의 디자인 프로세스를 혁신적으로 개선하고 새로운 비즈니스 기회를 창출하는 데에 기여하고 있다[6, 7]. 이러한 상황에서 의류 도면 분야에서도 디지털화 및 디지털 트랜스포메이션에 대한 적극적인 도입이 필요한 상황이다.

디지털화된 의류 디자인 도면은 다양한 산업과 연계할 수 있는 잠재력을 가지고 있다. 예를 들어, 도면 이미지만 확보하더라도 이미지 내 글자를 검출하고 인식하여 의류의 치수 정보를 바탕으로 디지털 아티스트나 가상 모델링 소프트웨어를 활용하여 의류를 가상 모델에 입혀보는 등의 시뮬레이션 작업을 수행할 수 있다[6, 8]. 이를 통해 제품 개발 과정에서 디자인을 더욱 효과적으로 평가하고, 소비자들에게 시각적으로 제품을 더 잘 전달할 수 있다. 따라서 빠르고 정확한 도면 인식 기술은 의류 디자인 도면을 디지털화하고 다양한 산업 군에 적용하기 위해서 반드시 필요한 기술이다.

최근 딥러닝을 기반으로 하는 기술들이 컴퓨터 비전 분야에서 높은 성능을 보여주고 있다. 특히, 딥러닝을 활용한 글자 검출과 인식 기술은 이미지 처리 분야에서 중요한 역할을 하며, 다양한 응용 분야에서 활용되고 있는 추세이다[9-11].

최근에 네이버 클로바 연구소에서 발표한 글자 검출 및 인식 딥러닝 모델은 높은 정확도와 성능을 보였으며, 이미지 기반 글자 검출 및 인식에서 좋은 결과를 보여주었다[12]. 그러나 의류 설계 도면 이미지는 일반적인 이미지와는 다른 고유한 특성을 가지고 있기 때문에, 해당 모델이

의류 설계도면 이미지에서도 뛰어난 성능을 발휘할 수 있는지에 대한 추가적인 평가가 필요하다. 또한 도면 이미지 내의 글자의 크기는 작지만 형태가 단순하기 때문에, 딥러닝 모델을 경량화하고 최적화하는 연구도 필요하다.

도면 내에서의 글자 이미지는 자연 영상에서의 문자 인식과 다르게 대체적으로 배경과 전경을 분리하기에 용이하다. 즉, 배경이 되는 도면은 거의 흰색이며 글씨는 검정, 빨강 등 한정된 색을 사용한다. 따라서 도면 내에서 글씨의 위치만 정확하게 찾아낼 수 있다면 문자 인식 부분에 있어서는 공개된 모델들을 사용하여도 충분히 높은 성능을 기대할 수 있다. 따라서 도면 인식에서 중요한 부분은 글씨의 위치를 정확하게 검출하는 것이며, 본 논문에서는 이미지 내에 있는 글자를 검출하는 딥러닝 모델 설계에 중점을 두었다.

추가적으로, 자연 영상 이미지에 비하여 비교적 단순한 의류 디자인 도면의 형태를 고려하여 경량화된 딥러닝 네트워크를 제안하였다. 또한 의류산업 특성상 많은 양의 의류 디자인 도면 이미지들을 확보하기 어렵기 때문에 적은 양의 데이터로도 성능을 보장할 수 있는 전이 학습 방법을 통하여 딥러닝 모델을 학습하였다.

마지막으로 [12]에서 제안한 글자 검출 딥러닝 모델과 본 논문에서 제안한 경량화된 딥러닝 모델에 대해 성능을 비교 평가하였다. 실험 결과, 제안한 딥러닝 모델이 의류 설계 도면 이미지에서의 문자 검출에 있어 약 20% 높은 성능을 보였다.

II. Related Works

의류 디자인 도면 이미지의 글자 검출과 인식을 위해 다양한 기술과 방법들이 연구되고 있으며, 이들 연구들은 최근 딥러닝의 발전과 함께 빠르게 진화하고 있다. 다음은 이미지 내 글자 검출에 관련된 주요한 연구들과 특징에 대해 기술한다.

1. Text Detection based on Image Processing

허프 변환과 컨투어 검출은 이미지에서 글자를 검출하는 데 사용되는 전통적인 알고리즘이다. 허프 변환은 이미지에서 선과 원을 검출하기 위해 활용되며, 컨투어 검출은 이미지에서 윤곽선을 검출하는 데에 사용되고 있다. 이러한 알고리즘들은 다양한 산업 군에서 문자 인식을 위한 관심 영역 도출을 위해 활용된다. 차량의 번호를 인식하기 위해 이미지에서 자동차의 차량 번호 영역을 검출하기 위

한 전처리 알고리즘으로 사용하거나[13, 14], 산업용 명판의 글자들을 인식하기 위해 이미지 내에서 명판의 위치를 찾기 위해서도 사용하고 있다[15]. 이미지 내 글자는 주변과 색깔 차이가 나고, 연속된 픽셀을 가진다는 특징을 가지기 때문에 윤곽선 검출을 통하여 관심 영역들을 그룹 지을 수 있다. 그리고 응용 애플리케이션의 요구 조건에 따라 허프 변환을 통하여 불필요한 선들을 제거하면 관심 영역들을 보다 세분화하여 글자 인식이 가능하도록 처리할 수도 있다.

영상처리 기반의 글자 검출 알고리즘들은 글자의 크기 및 형태가 정해져 있거나 글자 위치의 변화가 크지 않을 경우에 효과적이다. 하지만 빛 반사 등과 같은 영상 취득 환경에 따른 이미지의 노이즈에 민감하기 때문에 복잡한 이미지 전처리 과정이 필요하다. 또한 의류 디자인 도면에 포함된 표와 같이 문자 이외의 배경도 글자와 같은 특성을 지니고 있거나 하나의 문자만 존재하는 경우, 전처리의 임계 값에 따라 글자 영역을 제대로 검출 못하는 한계점이 있다.

2. Text Detection based on Deep Learning Scheme

최근에는 딥러닝을 기반으로 하는 글자 검출과 인식 기술들이 주목받고 있으며[16, 17], 이미지 처리 분야에서 높은 성능을 보여주고 있다. 딥러닝은 이미지에서 글자를 검출하고 인식하는데 특화된 모델을 학습하여 정확한 글자 검출과 인식을 수행할 수 있고, 허프 변환과 컨투어 검출의 한계를 극복하고 이미지의 노이즈에도 강한 강점을 가지고 있다. 딥러닝 기반의 글자 검출 모델을 학습하기 위해서는 ICDAR (International Conference on Document Analysis and Recognition), MSRA-TD500 (MSRA Text Detection 500)에서 제공하는 오픈 데이터 셋을 많이 활용하고 있으며, CNN(Convolution Neural Network), RNN(Recurrent Neural Network), LSTM(Long Short Term Memory) 등과 같은 네트워크들을 조합하여 딥러닝 네트워크를 설계하고 모델을 학습한다.

2019년 Baek et al.[12]에 의해 제안된 딥러닝 기반의 텍스트 검출 알고리즘인 CRAFT(Character-Region Awareness For Text detection)는 효과적이고 정확한 텍스트 영역 검출을 위해 개발되었다. 기존 딥러닝 기반의 문자 검출에 사용되는 공개 데이터 셋의 GT(Ground Truth)가 단어 수준(word-level)으로 구성되어 있어, 문자의 배치가 다양한 형태로 변형된 경우, 문자 검출에 대한 성능이 저하되는 문제점이 있었다. CRAFT는 이런 문제를 해결하면서도 높은 정확도와 속도를 동시에 제공하

는 특징을 갖고 있으며, 이를 달성하기 위해 텍스트 영역을 개별 문자 수준(Character-level)으로 인식하고, 이를 텍스트 영역으로 병합하여 검출하는 접근 방식을 채택하였다. 이러한 개별 문자 단위의 인식은 텍스트 영역의 규칙적이고 구조화된 특성을 이용하여 더욱 정확한 검출을 가능케 하였다. 또한 텍스트 영역을 찾는 데 있어 어려운 문제를 효과적으로 다루는 다양한 기술들을 적용하였으며, 특히, 텍스트 영역과 배경을 잘 구분할 수 있는 새로운 손실 함수를 도입하고, 멀티스케일 특징 추출을 통해 다양한 텍스트 크기에 대응하도록 설계되었다. 또한, 이미지 내에서 텍스트가 형성하는 특정 패턴과 구조를 인식하도록 학습하였으며, 이를 통해 정확하게 텍스트 영역을 검출하는 성능을 향상시켰다.

CRAFT의 성능은 ICDAR 및 MSRA-TD500에서 제공하는 공개 데이터 셋을 활용하여 Recall, Precision, F1-Score 항목들을 평가 지표로 사용하여 검증하였다. CRAFT는 MSRA-TD500에서 Recall, Precision, F1-Score 각각에 대하여 78.2%, 88.2%, 82.9%의 성능을 보였으며, [17]에서 제안한 텍스트 검출 알고리즘은 76.2%, 87.6%, 81.5%의 성능을 보였다. ICDAR 및 MSRA-TD500 데이터 셋에 대해서 CRAFT가 [16, 17]에서 제안한 알고리즘들보다 약 2%의 높은 성능을 보여주었다[12].

이러한 CRAFT의 텍스트 영역 검출 성능과 효율성은 의류 디자인 도면 이미지 분석 분야에서도 좋은 성능을 보여 줄 것으로 기대하고 있으며, 문자 단위 인식과 멀티스케일 특징 추출 기법이 이러한 이미지에서의 글자 검출에 높은 성능을 발휘할 것으로 기대된다. 하지만 CRAFT는 비용 문제 등을 이유로 개별 문자 단위(Character-Level)로 된 데이터 셋을 활용하여 모델을 만들지 않고 단어 단위(Word-Level)를 개별 문자 단위로 추정하는 문제로 학습하였기 때문에 개별 문자 단위에 대해서는 정확하게 검출 못 할 수 있다. 의류 도면 이미지에는 문자가 하나만 있는 경우가 많기 때문에 의류 도면 이미지에서 문자 검출을 하기 위해서는 CRAFT 모델을 그대로 사용할 수 없다.

3. Text Recognition

문자 인식 기술은 검출된 문자 영역이 어떤 글자인지를 인식하는 기술이다. 문자 인식 기술에 있어서도 딥러닝 기반의 기술들이 높은 인식률을 보여주며 SOTA(State of The Art)로 인정받고 있다. 국내외의 다양한 기업 및 연구에서 딥러닝 기술을 기반으로 한 글자 인식 모델에 관한 연구를 진행 중이며, 오픈소스 진영에서도 Tesseract[18], EasyOCR[19] 등 활발한 연구가 진행 중이다.

한편, 문자 인식에 관한 연구가 활발해지면서 제안된 문자 인식 모델에 대한 객관적인 성능 평가의 필요성도 제기되었다. [20] 논문에서는 기존에 연구된 글자 인식 모델들의 객관적인 성능 평가를 위한 방법을 제안하였다. 동일한 테스트 세트를 구축하고 여러 딥러닝 모델들을 조합하여 정확도를 비교하였다. [20]에 의하면 TPS(Thin Plane Spline[21]) + ResNet[22] + BiLSTM[23] + CTC(Connectionist Temporal Classification[24]) 방법이 인식을 면에서 가장 우수한 성능을 보였다.

본 논문에서는 문자 인식을 위하여 EasyOCR의 문자 인식 알고리즘을 사용하였다. EasyOCR은 문자 검출 방법으로는 네이버 클로바AI 팀의 CRAFT 모델[12]을 사용하고 있으며, 문자 인식을 위해서는 [20]에서 제안된 것과 유사한 ResNet + BiLSTM + CTC 방식을 사용하여 구현되어 있다. 본 논문은 의류 설계 도면에서 문자 검출의 성능을 높이기 위한 경량화 모델을 제안하는 연구이므로, 문자 인식 모델을 위한 별도의 학습은 진행하지 않았다.

III. The Proposed Scheme

이미지 내 글자 검출은 일반적으로 이미지에서 글자를 탐지하고 추출하는 작업을 의미하며, 의류 도면 이미지는 옷의 디자인과 치수 정보 등을 포함하고 있는 특수한 종류의 이미지이다. 의류 도면 이미지 내 글자 검출은 옷의 디자인과 치수를 이해하는 데 매우 중요하며, 더 나아가 3D 모델 피팅과 같은 응용에 필수적인 기초 자료로 활용된다. 관련 연구에서 언급한 [12] 모델은 이미지 내 글자 검출에 사용할 수 있는 효과적인 모델 중 하나이지만 의류 도면 이미지와 같이 도면 내 표 안의 내용들이 문자 하나만 존재하거나 작은 숫자들이 포함된 경우, 글자 검출의 성능이 저하되는 문제점이 발생할 수 있다.

그래서 본 논문에서는 위에서 언급한 의류 도면 이미지의 고유한 특성을 고려하여 별도로 수집한 의류 도면 이미지 내 문자들을 어노테이션하여 약 700장의 데이터 셋을 구축하였다. 또한 의류 도면 이미지의 형태가 단순하기 때문에 CRAFT 모델을 기반으로 한 경량화된 딥러닝 네트워크를 설계하였다. 그리고 산업 특성상 많은 의류 도면 이미지를 수집할 수 없어 적은 수량의 데이터를 활용하여 딥러닝 모델을 생성해야 하므로 본 논문에서는 전이학습 방법을 활용하였다. 제안한 경량화된 딥러닝 네트워크는 [12]에서 사용한 VGG-16 [25] 백본 네트워크의 학습 파라미터와 마지막 레이어를 사전 훈련된 가중치(pre-trained

weights)로 활용하였다.

이 방법을 통해 학습된 모델은 의류 도면 이미지에서 글자를 더욱 정확하게 검출할 수 있음을 실험을 통해 확인하였다. 특히 단일 문자 및 작은 수치를 정확하게 검출함으로써 미검출 성능을 평가하는 지표로 사용되는 Recall 성능이 약 29% 향상된 것을 확인하였다.

1. Dataset

문자 검출 딥러닝 모델 학습을 위해 [26]에서 사용하는 오픈 데이터 셋을 기본적으로 사용하였으며, 의류 도면의 글자 검출 성능 향상을 위해 별도 수집된 패션 도면 이미지를 어노테이션하고 학습 데이터 셋에 추가하였다.

총 704장의 패션 도면 이미지를 어노테이션하였으며, 그중 679장은 학습 데이터 셋으로 사용하여 네트워크를 학습시켰다. 나머지 25장 중 15장은 학습 중 Validation을 위한 목적으로 사용하였으며, 10장은 모델의 성능 검증을 위해 사용하였다.

Fig. 1.은 데이터 셋 및 어노테이션 예시를 보여주고 있으며, 어노테이션을 위해 [27]에서 제공하는 툴을 사용하였다. 문자 검출을 위해 “작성일자”라는 단어를 Bbox (Bounding Box) 형태로 저장하고 있으며, 사각형의 각 꼭짓점에 대한 이미지 좌표들이 저장된다. 추가적으로 추후 문자 인식을 위해 해당 글자가 무엇인지에 대한 정답 값 (“작성일자”)도 같이 저장하였다.

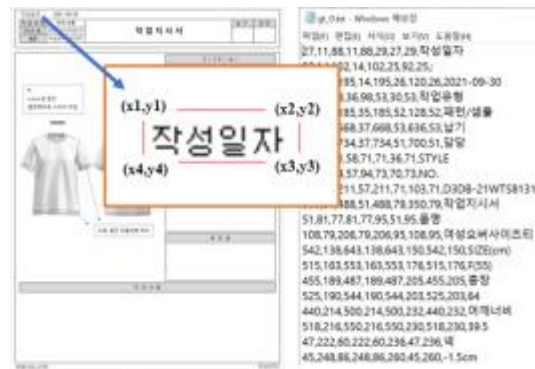


Fig. 1. Example Dataset and Annotations

2. Proposal Architecture

Fig. 2.는 본 논문에서 제안한 경량화 딥러닝 네트워크 아키텍처를 보여주고 있다. Conv.는 Convolution Layer를 의미하며, h, w는 각각 height, width를 나타낸 것이다. 그리고 괄호 내의 숫자(예: (h/w, w/2, 64))는 Convolution 연산 수행 후의 출력 이미지 사이즈 및 채널을 표현하였다. UpConv.는 이미지 사이즈를 키우는

Up-Sampling을 통해 출력된 이미지에서 Convolution을 수행하는 것을 의미한다.

백본 네트워크는 [12]에서 사용한 VGG-16을 사용하였고, 제안한 모델의 디코딩 부분에는 스킵 연결이라고 불리는 중요한 구성 요소를 포함하였다. 스킵 연결은 저수준 특징 정보를 고수준 특징 정보와 결합하는 데 사용되며 [28], 이것은 모델이 이미지의 세부 정보를 보존하고 텍스트 영역을 정확하게 검출하는 데 도움을 준다. 그리고 경량화를 위하여 각 UpConv. Layer들의 채널 수를 반으로 줄였으며, 마지막 두 개의 Layer들을 제거하고 채널 수를 조정하여 학습하였다.

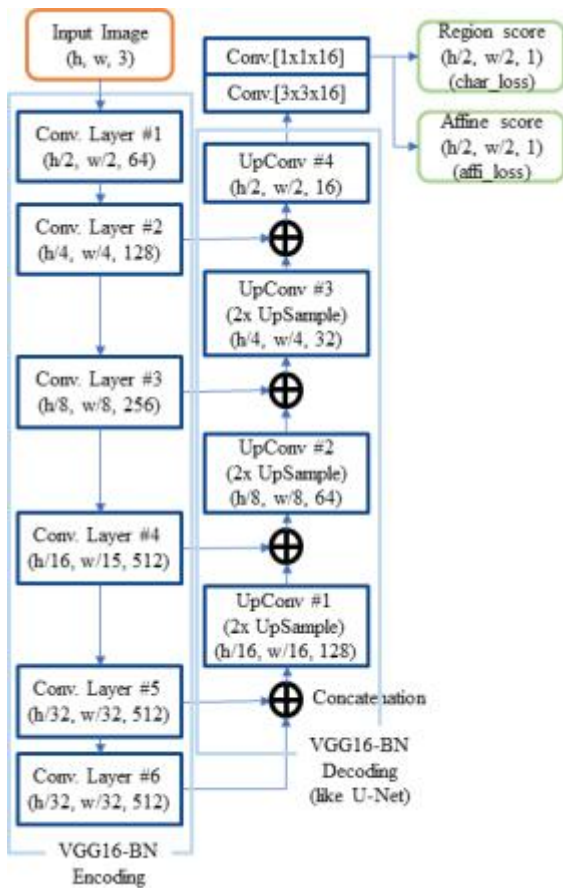


Fig. 2. Proposal Network Architecture

모델의 최종 출력은 두 가지 주요 채널을 가지고 있다. 첫 번째 채널은 Region Score라고 불리며, 이것은 각 픽셀이 텍스트 영역의 일부인 확률을 나타낸다. 두 번째 채널은 Affine Score라고 불리며, 이것은 두 픽셀 간의 유사성을 측정한다. 본 논문에서는 텍스트 영역(Region)에 대한 손실인 {char loss}와 글자 위치 및 형태에 대한 손실을 표현하는 {affi loss}을 결합하여 전체 Loss Function인 {total loss}를 Eq. 1.과 같이 정의하였다.

{char loss}는 Eq. 2.와 같이 정의하였고, {affi loss}는 Eq. 3.과 같이 정의하였으며, 두 loss 모두 MSE(Mean Square Error)를 사용하여 계산하였다.

$$\{total\ loss\} = \{char\ loss\} + \{affi\ loss\} \quad (Eq. 1)$$

$$\{char\ loss\} = \frac{1}{N} \sum_{i=1}^N (\alpha \cdot \{MSE\}(p_i, y_i) + (1 - \alpha) \cdot \{batch\ image\ loss\}(p_i, y_i)) \quad (Eq. 2)$$

$$\{affi\ loss\} = \frac{1}{N} \sum_{i=1}^N (\alpha \cdot \{MSE\}(q_i, z_i) + (1 - \alpha) \cdot \{batch\ image\ loss\}(q_i, z_i)) \quad (Eq. 3)$$

char loss 및 affi loss는 개별 픽셀 수준에서 계산하는 방법이며, α 는 두 손실 간의 상대적인 중요성을 조절하며, batch image loss에서 글자 영역과 글자 영역이 아닌 배경 부분에 대한 손실을 계산하여 가중 평균을 구하였다.

제안한 경량화된 딥러닝 모델을 생성하기 위해, 학습 시 batch_image_loss에서 사용되는 글자 영역과 아닌 영역의 픽셀 비율을 0.3으로 세팅하였으며, 글자 영역이 아닌 최소 픽셀 수는 5,000으로 세팅하였다. 그리고 최적화를 위해 Adam optimizer를 사용하였으며, epoch은 24,000, learning rate는 1e-4로 세팅하였다. NVIDIA Quadro RTX 5000 2개가 장착 되어있는 워크스테이션으로 학습을 진행하였으며, 약 2,900분 정도 소요되었다.

Fig. 3.은 제안한 딥러닝 모델과 기존 방법[12]의 검출 결과를 비교한 예시이다. Fig. 3. a)에서 보듯이 기존 방법을 사용하였을 경우에는 한 단어로 인식되어야 하는 단어들이 각각의 영역으로 검출되기도 하고 (예: 뒷+길이, 화+장, 네크라인립단+폭) 각각 인식되어야 하는 치수들이 하나의 텍스트 영역으로 인식되기도 하였다(Fig. 3. D). 그러나 제안한 방법에서는 음절의 분리 없이 하나의 단어 영역으로 검출되며, 치수 또한 정확하게 인식하는 것을 확인할 수 있었다. Fig. 4. 는 제안한 방법으로 개선된 문자 영역 검출의 또 다른 예시이다. 상기 기술한 것처럼 [12] 방법에서는 낱자를 검출에 있어 검출률이 떨어지는 단점이 있으며, 의류 설계 도면에서는 알파벳 한 글자로 항목 번호를 표기하는 경우가 많아 검출률이 낮아지게 된다. 그러나 제안한 방법에서는 이와 같은 단점을 보완하여 학습을 진행하였기 때문에 낱자 검출에서 우수한 검출 성능을 보였다. 전체 테스트 데이터 세트에 대한 성능 평가는 4장에서 기술하도록 한다.

뒷길이	43343	3143
가슴놀래	4343	43
어깨너비	75	3
밑단놀래	43	23
소매길이	34	221
화장	43	33
소매동	86	34
소매단놀래	43	54
목폭	6575	53
목길이	567	43
암홀(직선)	765	54
소매립단폭	56	64
밑단립단폭	567	657
네크라인립단폭	567	567
	567	845

a) Text Detection Result of [12]

뒷길이	43343	3143
가슴놀래	4343	43
어깨너비	75	3
밑단놀래	43	23
소매길이	34	221
화장	43	33
소매동	86	34
소매단놀래	43	54
목폭	6575	53
목길이	567	43
암홀(직선)	765	54
소매립단폭	56	64
밑단립단폭	567	657
네크라인립단폭	567	567
	567	845

b) Text Detection Result of our model

Fig. 3. Results of proposal our model inference

I	앞목길이	457
J	뒤목길이	57
K	암홀	456
L	네크라인놀이(립)	456
M	소매단놀이(립)	456
N	밑단놀이(립)	56
O		
P		
Q		

I	앞목길이	457
J	뒤목길이	57
K	암홀	456
L	네크라인놀이(립)	456
M	소매단놀이(립)	456
N	밑단놀이(립)	56
O		
P		
Q		

a) Text Detection Result of [12]

b) Text Detection Result of our model

Fig. 4. Different Results of Text Detection

IV. Experiment

본 논문에서 제안한 문자 검출 경량화 딥러닝 모델과 CRAFT 모델을 추가 수집한 패션 도면 이미지 중 테스트 데이터 셋 10장에 대해 성능을 비교 평가하였다. 성능 평가는 정밀도(precision), 재현율(recall), 그리고 조화평균(F1-score)을 계산하여 진행하였으며, 모델의 성능을 GUI

툴을 통해 시각적으로도 확인하였다.

$$Precision = \frac{TP (True Positive)}{TP (True Positive) + FP (False Positive)} \quad (Eq. 4)$$

$$Recall = \frac{TP (True Positive)}{TP (True Positive) + FN (False Negative)} \quad (Eq. 5)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (Eq. 6)$$

- 정밀도(precision): Eq. 4와 같이 모델이 TP 클래스로 예측한 샘플 중에서 실제로 TP 클래스인 샘플의 비율을 나타내는 평가 지표로, FP, 즉, 오검출을 반영하는 지표이다.

- 재현율(recall): Eq.5와 같이 실제 TP 클래스인 샘플 중에서 모델이 정확하게 TP 클래스로 예측한 샘플의 비율을 나타내는 매트릭스로, 미검출을 반영하는 지표로써, FN을 줄이는 데 중점을 둔다.

- 조화평균(F1-score): Eq. 6과 같이 정밀도와 재현율의 조화 평균으로, 정확도와 재현율 사이의 균형을 평가하여 모델의 성능을 종합적으로 이해하는 데 도움을 준다.

Table 1은 CRAFT와 제안한 딥러닝 모델에 대한 성능 평가 결과를 나타낸 것이다.

Table 1. Model Accuracy Evaluation results

Metrics	Network		
	CRAFT	CRAFT	Proposed
	Dataset		
	SynthText + ICDAR15	Ours	Ours
Precision	0.88	0.97	0.99
Recall	0.64	0.95	0.93
F1-Score	0.74	0.96	0.96

Fig. 3. 결과와 같이 CRAFT에서는 문자 하나만 있거나 숫자만 있는 경우, 정확하게 검출하지 못하거나 위/아래의 내용이 합쳐서 검출하는 경우가 있었다. 특히, 문자 하나만 있는 경우에 대해 정확하게 검출 못한 문제로 인하여 Recall 성능이 실험 결과 약 64%로 나타났다.

CRAFT 모델을 기반으로 별도 제작한 Dataset을 추가 하여 학습한 모델의 성능이 가장 좋은 것으로 확인되었다. 하지만 본 논문에서 제안한 방식으로 딥러닝 네트워크를 경량화하더라도 약 0.5% 성능밖에 저하되지 않는 것을 시험을 통하여 확인하였다.

Table 2는 제안한 경량화 딥러닝 모델과 기존 모델과의 파라미터 개수 및 처리 속도에 대해 비교 평가한 결과이다.

Table 2. Comparative evaluation of the number of parameters and Model Inference time

Metrics	CRAFT	Proposed Network
Networks parameters	20,770,466	19,717,346
Model Size	243 MB	231 MB
Average Inference Time	0.70 Sec.	0.66 Sec.

제안한 경량화된 모델이 기존 딥러닝 모델보다 파라미터 수와 모델 사이즈 모두 줄어든 것을 확인하였으며, 딥러닝 모델 추론 속도도 평균 40ms 향상되었다.

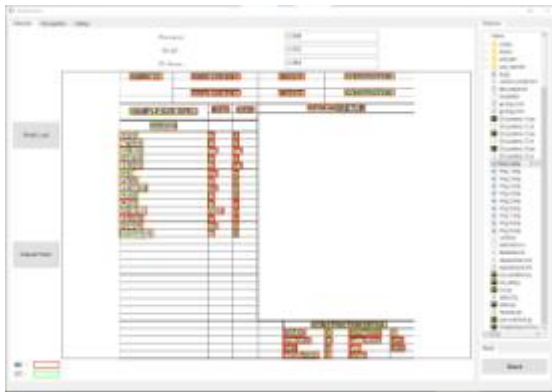


Fig. 5. Result of text detection on the GUI tool

본 연구에서는 제안한 모델의 검출 결과를 용이하게 확인하기 위하여 Python 기반의 GUI SW를 개발하였다. 개발한 GUI SW에서 학습된 모델을 선택하고, 도면 데이터를 선택하면 모델의 검출 결과를 화면에 보여준다.

Fig. 5.는 개발한 GUI 툴 상에서 글자 검출 결과를 보여준다. 초록색은 GT(Ground Truth)를 표현한 것이고, 빨간색 Bbox(Bounding Box)가 본 논문에서 제안한 모델로 검출한 결과이다. GUI 툴을 통하여 각 샘플마다 글자 검출이 잘 되는지를 시각화하여 결과를 확인할 수 있어 성능 평가에 보다 신뢰성을 높였다.

V. Conclusions

본 논문에서는 의류 디자인 도면 이미지의 글자 검출을 위한 경량화된 딥러닝 모델을 제안하였다. 기존의 딥러닝 모델 중 글자 검출 성능이 뛰어난 모델을 선택하고, 의류 도면 이미지의 특수성을 고려하여 모델을 최적화하였다. 우리가 제안한 딥러닝 네트워크는 VGG16 백본 네트워크를 기반으로 하되, Decoding 부분과 마지막 Layer를 경

량화하는 방향으로 설계하였다. 또한, 의류 도면 데이터셋을 수집하고 이를 활용하여 모델을 학습시켰으며, 실험 결과, 제안한 네트워크는 기존 딥러닝 기반의 글자 검출 모델보다 약 20% 더 높은 성능을 보였다. 이러한 결과는 의류 디자인 도면 이미지 분야에서 글자 검출의 효율성을 향상시킬 수 있을 것으로 기대된다.

향후 연구 방향으로 높은 성능을 위해 딥러닝 네트워크의 추가적인 최적화를 진행할 예정이며, 의류 디자인 도면 이미지의 특성을 보다 정교하게 고려한 모델 개발이 중요한 과제이기 때문에 특수한 글자 정보를 고려한 모델 성능 향상 방안을 추가 연구하고자 한다.

의류 디자인 도면 이미지의 디지털 트랜스포메이션은 의류 산업에 혁신과 생산성 향상을 가져올 것으로 예상되며, 딥러닝 기술의 발전된 활용을 통해 의류 디자인 분야에서 새로운 비즈니스 기회를 창출하는데 기여할 것으로 기대된다.

ACKNOWLEDGEMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(No.1711193917, Development of XR Collaborative Platform Based on Digital Fashion Asset).

REFERENCES

- [1] B. Rathore, "Fashion Transformation 4.0 : Beyond Digitalization & Marketing in Fashion Industry," *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal*, Vol. 10, No. 2, pp. 54-59, 2021. DOI: <https://doi.org/10.56614/eiprmj.v10i2.234>
- [2] O. Behr, "Fashion 4.0 - Digital Innovation in the Fashion Industry," *Journal of technology and innovation management*, Vol. 2, No. 1 pp. 1-9, 2018. DOI: 10.1108/RJTA-03-2018-0023
- [3] H. J. Choo, Y. J. Nam, Y. R. Lee, H. K. Lee, S. J. Lee, S. E. Lee, J. I. Jang, J. H. Park, J. W. Choi, and D. Y. Kim. "Domestic Research Trends in IT Fashion," *Journal of the Korean Society for Clothing Industry*, Vol.14, No. 4, pp. 614-628, August, 2012. doi:10.5805/ksci.2012.14.4.614
- [4] K. S. Noh, "A study on Strategic Fitness of Digital Transformation and Competitive Strategy," *Journal of the Korean Convergence Society*, Vol.12, No. 8, pp. 257-265, August, 2021.

- [5] B. Vial, "Understanding digital transformation: A review and a research agenda," *Managing Digital Transformation*, pp. 54, 2021.
- [6] Chamorro-Premuzic, T., "The essential components of digital transformation," *Harvard Business Review*, Vol. 13, pp. 1-6, 2021.
- [7] S.V. Akram, P.K. Malik, R. Singh, A. Gehlot, A. Juyal, K. Z. Ghafoor and S. Shrestha, "Implementation of Digitalized Technologies for Fashion Industry 4.0: Opportunities and Challenges", *Scientific Programming*, Vol. 2022, 17 pages, 2022. <https://doi.org/10.1155/2022/7523246>
- [8] R. Choi, and C. S. Cho. "A Virtual Fitting System Using The Top and Bottom Image of Garment." *Journal of Korea Multimedia Society*, Vol. 15, No. 7, pp. 941-950, July 2012. DOI:10.9717/kmms.2012.15.7.941
- [9] L. H. Yang, P. Zhang, F., "Review of Scene Text Detection and Recognition," *Journal of Arch Computat Methods Eng*, Vol. 27, pp. 433-454, 2020. DOI: <https://doi.org/10.1007/s11831-019-09315-1>
- [10] S. N. Matton, A. Greaves, M, and Lam, A. "A survey of deep learning approaches for ocr and document understanding," arXiv preprint arXiv:2011.13534, 2020.
- [11] M. Wang, S. Niu, and Z. Gao, "A novel scene text recognition method based on deep learning," *Journal of Computers, Materials & Continua*, Vol. 60, No. 2, pp. 781-794, 2019.
- [12] J. Y. Baek, J. Y. Lee, B. Han, T. K. Kim, and K. H. Lee, , "Character Region Awareness for Text Detection," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10507-10516, 2019.
- [13] V. Shapiro, D. Dimov, S. Bonchev, V. Velichkov and G. Gluhchev, "Adaptive license plate image extraction," *International Conference on Computer Systems and Technologies*, 2003.
- [14] S. Kim, D. Kim, Y. Ryu and G. Kim, "A Robust License-plate Execution Method under Complex Image Conditions," the 16th International Conference on Pattern Recognition (16th ICPR), Quebec, Canada, pp. 216-219, 2002.
- [15] H. Li, Y. Ma, H. Bao, and Y. Zhang, "Probabilistic Hough Transform for Rectifying Industrial Nameplate Images: A Novel Strategy for Improved Text Detection and Precision in Difficult Environments," *Journal of Appl. Sci*. Vol. 13, No. 7, pp. 4533. DOI: <https://doi.org/10.3390/app13074533>
- [16] P. Lyu, M. Liao, C. Yao, W. Wu and X. Bai, "Mask textspotter: An end-to-end trainable neural network for spotting text with arbitrary shapes," arXiv preprint arXiv:1807.02242, 2018.
- [17] M. Liao, B. Shi, and X. Bai, "Textboxes++: A single-shot oriented scene text detector," *Image Processing*, Vol. 27, No. 8, pp. 3676-3690, 2018.
- [18] R. Smith, tesseract, <https://github.com/tesseract-ocr/tesseract>
- [19] Jaided AI, EasyOCR, <https://github.com/JaidedAI/EasyOCR>
- [20] J. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S. Oh and H. Lee, "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis," *International Conference on Computer Vision*, 2019.
- [21] M. Jaderberg, K. Simonyan, A. Zisserman and K. Kavukcuoglu, "Spatial Transformer Networks," *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pp. 2017-2025, Dec. 2015.
- [22] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016. DOI: 10.1109/CVPR.2016.90.
- [23] B. Shi, X. Wang, P. Lyu, C. Yao and X. Bai, "Robust Scene Text Recognition with Automatic Rectification," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 4168-4176, 2016. DOI: 10.1109/CVPR.2016.452.
- [24] A. Graves, S. Fernandez, F. Gomez and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," *Proceedings of the 23rd international conference on Machine learning*, June, pp. 369-376, 2006. DOI: <https://doi.org/10.1145/1143844.1143891>
- [25] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, pp. 1097-1105, 2014.
- [26] Z. Wang, J. Zhang, and Z. Wang, "EasyOCR: A fast, accurate, and user-friendly OCR system," arXiv preprint arXiv:2109.13099. DOI: 10.1109/ICASSP.2022.9731104
- [27] wkentaro, labelme, <https://github.com/wkentaro/labelme>
- [28] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation," In *MICCAI 2015*, pp. 234-241, 2015.

Authors



Ju-Seok Shin received the B.S. degree in Computer Science from Kyungil University, Korea in 2006. Ju Seok Shin received the M.S. and Ph.D. degrees in Computer Science from Kyungpook National University, Daegu,

Korea in 2010 and 2014, respectively. From 2010 to 2014, he worked as a Senior Engineer in the Electronics and Telecommunications Research Institute. During 2014-2021, he worked as a chief researcher in the Delphi autonomous driving system development team. He is currently a Professor in the Department of Artificial Intelligence at Korea Polytechnics. He is interested in autonomous driving system, embedded system, machine learning and deep learning and computer vision system.



Hyun-Woo Kang received the B.S., M.S. and Ph.D. degrees in Computer Science from Kyungpook National University, Daegu, Korea, in 2005, 2009 and 2017, respectively. From 2009 to 2017, he worked as a Senior

Engineer in the Electronics and Telecommunications Research Institute. He is currently a Professor in the Department of Artificial Intelligence at Korea Polytechnics. He is interested in wireless network, machine learning and deep learning and computer vision system.