

변형 VGG 모델의 전처리를 이용한 부품도면 문자 인식 성능 개선

신희란* · 이상협* · 박장식** · 송종관*

Performance Improvement of Optical Character Recognition for Parts Book Using Pre-processing of Modified VGG Model

Hee-Ran Shin* · Sang-Hyeop Lee* · Jang-Sik Park** · Jong-Kwan Song*

요 약

본 논문에서는 기계 서비스 부품 도면에서 숫자를 인식하기 위하여 입력 영상에 대한 전처리와 딥러닝 모델을 제안한다. 서비스 부품 도면의 숫자를 인식하는데 있는 지시선과 도형에 의한 오검출 또는 오인식을 개선하기 위하여 수학적 형태학 필터링 전처리를 한다. 숫자 인식을 위하여 VGG-16 모델을 축소 변형한 7 개의 계층을 가지는 VGG 모델을 적용함으로써 인식 성능을 개선한다. 서비스 부품 도면의 숫자 인식 실험 결과, 제안하는 방법이 인식률 95.57%, 정확도는 92.82%로 종래의 방법에 현저히 개선된 결과를 얻었다.

ABSTRACT

This paper proposes a method of improving deep learning based numbers and characters recognition performance on parts of drawing through image preprocessing. The proposed character recognition system consists of image preprocessing and 7 layer deep learning model. Mathematical morphological filtering is used as preprocessing to remove the lines and shapes which causes false recognition of numbers and characters on parts drawing. Further.. Further, the used deep learning model is a 7 layer deep learning model instead of VGG-16 model. As a result of the proposed OCR method, the recognition rate of characters is 92.57% and the precision is 92.82%.

키워드

Deep Learning, Mathematical Morphology Filtering, Optical Character Recognition,
광학 문자 인식, 심층 학습, 수학적 형태학 필터

1. 서 론

광학 문자 인식(Optical character recognition; OCR)은 사람이 작성하거나 인쇄된 문자를 컴퓨터로 편집할 수 있는 문자 코드 형식으로 변환 기술이다.

OCR은 차량 번호판 인식, 금융 거래의 신분증 및 카드 번호 인식, 금융 및 의료 서류 워드 변환 등의 다양한 산업 분야에서 적용되고[1-3]. 최근에는 인공지능 번역에도 적용되어 문자를 인식하여 번역을 하는 서비스 또는 웹사이트 검색하는 서비스, 연관 데이터

* 경성대학교 전자공학과(heeran3646@gmail.com, * Received : Mar. 07, 2019, Revised : Mar. 26, 2019, Accepted : Apr. 15, 2019

,jsipark@ks.ac.kr, jsong@ks.ac.kr)

• Corresponding Author : Jang-Sik Park

** 교신저자 : 경성대학교 전자공학과

Dept. Department of Electronic Engineering, Kyung Sung University,

• 접수일 : 2019. 03. 07

Email : jsipark@ks.ac.kr

• 수정완료일 : 2019. 03. 26

• 게재확정일 : 2019. 04. 15

검색 등 다양한 부분에서 OCR이 응용되고 있다[4].

기계부품 산업에서는 부품 도면(Parts book)을 제작할 때, 설계 부품도면과 서비스 부품도면 두 종류를 만든다. 설계 부품도면에는 부품의 상세한 정보를 포함하기 때문에 정보 유출을 막기 위해 설계 부품도면을 배포하지 않고, 서비스 부품도면을 별도로 제작하여 배포한다. 서비스 부품도면의 제작은 일러스트레이터(Illustrator)와 같은 그래픽 도구를 사용한다. 각 부품에 대한 명칭과 설명을 위하여 서비스 부품도면에 각 부품에는 숫자와 문자를 연결한다. 도면의 숫자 또는 문자를 부품 명칭과 연결은 사람이 수작업으로 수행하고 있어 시간과 비용이 소요되고 있는 실정이다. 서비스 부품도면을 효율적으로 관리하기 위하여 도면의 숫자 또는 문자를 자동으로 인식하는 것이 필요하다. 서비스 부품도면의 숫자를 인식하기 위하여 일반적인 문서의 문자 또는 숫자 인식에 널리 사용되는 OCR 오픈 소스(Open source) Tesseract OCR[5]을 적용하였으나, 일반 문서에 대해서는 높은 인식률을 나타내지만, 서비스 부품 도면의 숫자 인식률은 현저히 낮다.

본 논문에서는 전처리와 딥러닝을 기반으로 서비스 부품도면의 숫자와 문자 성능 향상 방법을 제안한다. 제안하는 부품도면 OCR의 구성은 전처리와 딥러닝 모델로 구성된다. 서비스 부품도면의 숫자와 문자에 대한 오검출 또는 오인식의 원인이 되는 지시선과 도형을 제거하여 인식률과 정확도를 향상시키기 위해 수학적 형태학 필터링(mathematical morphology filtering)을 전처리로 적용한다. 숫자 인식은 도면 숫자 크기를 고려하여 기존 VGG-16을 변형한 축소 VGG(Mini-VGG) 모델을 이용한다. 제안하는 방법으로 도면 100장에 대한 숫자인식 실험 결과, 인식률(Recall) 95.57%, 정확도(Precision)는 92.82%로 종래의 방법에 대하여 현저히 개선된 결과를 얻었다.

II. 축소 변형 VGG 모델

VGG-16 모델은 컨벌루션(Convolution), 풀링(Pooling), FC(Fully connected) 계층으로 구성되어 있으며 16 계층으로 이루어져 있다. 커널 크기를 3×3으로 반복 수행함으로써 커널 크기가 큰 컨벌루션과

대응할 수 있도록 한다[6-7].

VGG-16 모델을 축소 변형 모델은 그림 1과 같이 데이터 입력 크기가 기존의 224×224를 대신 28×28로 7 계층으로 구성된다. 커널 크기(Kernel size)는 3×3으로 VGG-16과 같다. 표 1은 축소 변형 VGG 모델을 상세히 나타낸 것이다.

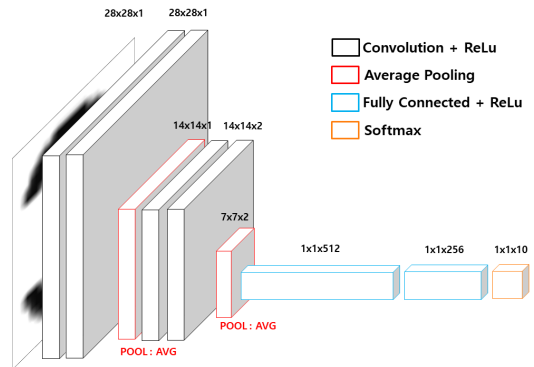


그림 1. 축소 변형 VGG 모델 구조
Fig. 1 Structure of mini-VGG model

표 1. 축소 변형 VGG 구조
Table 1. Mini VGG architecture

layer(type)		size
conv1_1	Convolution	28x28x1
conv1_2	Convolution	28x28x1
pool_1	Avg Pooling	14x14x1
conv2_1	Convolution	14x14x2
conv2_2	Convolution	14x14x2
pool_2	Avg Pooling	7x7x2
fc1	FC	512
fc2	FC	256
fc3	FC	10
softmax	FC	10

부품도면의 문자와 숫자 크기는 넓이 30부터 40까지, 높이 20에서 30까지 작은 크기의 이미지로써 학습 과정에서 이미지 손실을 줄이기 위해 입력 데이터의 크기가 작은 축소 변형 VGG가 적절하다. OCR 애플리케이션에 학습된 모델을 적용하여 숫자를 인식하는 과정에서도 추론 속도도 계층의 크기가 작기 때문에 축소 변형 VGG가 속도가 빠르다. 인식률 정확도와

처리 속도가 빠르기 때문에 축소 변형된 모델이 부품도면 OCR에 적절하다. 축소 변형 VGG 모델은 크기가 작은 객체에 대하여 기존 VGG-16보다 우수한 성능을 보인다.

III. 제안하는 전처리와 결합한 OCR

3.1 제안하는 OCR 구조

일반적 문서와 달리 지시선과 도형을 포함하는 서비스 부품도면의 OCR은 지시선이 '1' 또는 '7'로 문자로 오인식되거나 부품의 도형이 '0'으로 오인식되는 문제가 발생한다.

지시선과 도형을 제거하는 전처리 수행한 후에 딥러닝 추론 과정을 거친다. 그림 2는 제안하는 부품도면 OCR 처리 과정을 나타낸 것이다. 부품도면에서 숫자와 좌표정보를 기록한 실측정보(Ground Truth)를 파일(txt)로 기록한 후 실측정보(Ground Truth) 파일을 활용하여 딥러닝 학습 DB를 제작하고 학습한다. 부품도면의 지시선과 도형에 의한 오검출을 줄이기 위한 전처리를 통하여 지시선과 도형을 제거하고 학습 모델을 적용하여 검출 및 인식한다.

본 논문에서는 지시선과 도형을 제거하기 위한 전처리로 수학적 형태학 필터를 활용하고, 학습 및 추론을 위한 축소 변형 VGG 모델을 적용한다.

3.2 수학적 형태학 필터링 전처리

인식률을 향상시키기 위해 오검출의 원인이 되는 지시선과 도형을 제거하기 위하여 수학적 형태학 연산을 수행한다. 수학적 형태학 연산은 마스크(Mask)를 이용하여 객체를 확장 및 축소하여 객체의 형태를 변형하는 영상처리 연산이다. 기본적으로 침식(erosion), 팽창(dilation) 연산이 있고 이를 복합적으로 사용하여 객체를 변형 시킨다[8].

침식 연산은 객체의 각 픽셀에 마스크를 적용했을 때 마스크가 255값을 가지는 픽셀에 대해 입력 영상도 255값을 가져야 결과값 255가 된다. 대상 위치에서 한 픽셀이라도 0값을 가지게 되면 결과값이 0이 되기 때문에 객체가 축소되는 현상이 나타난다. 영상의 객체와 배경 사이에 잡음이 있을 때, 잡음을 제거하거나 임의의 작은 객체를 제거하는데 응용되며 최

솟값 필터의 역할을 수행하는 연산이다[9].

팽창 연산은 각 픽셀의 마스크를 적용했을 때 마스크의 유효 영역에 존재하는 픽셀들을 모두 255값으로 밝게 한다. 객체가 팽창하게 되는 현상이 나타난다. 영상의 내부 구멍이나 공간을 채우거나 짧게 끊어진 영역을 연결하는데 사용되며 최대값 필터의 역할을 하는 연산이다.

침식 연산과 팽창 연산을 수행하는 방법에서 열림 연산(opening)과 닫힘 연산(closing)있다. 열림 연산은 침식 연산을 수행한 후에 팽창 연산을 수행하여 객체 영역에서 세부 영역을 제거하는 연산이다. 닫힘 연산은 빈틈이 팽창 연산을 수행한 후에 침식 연산을 수행하여 객체 사이의 빈틈을 채운다[10].

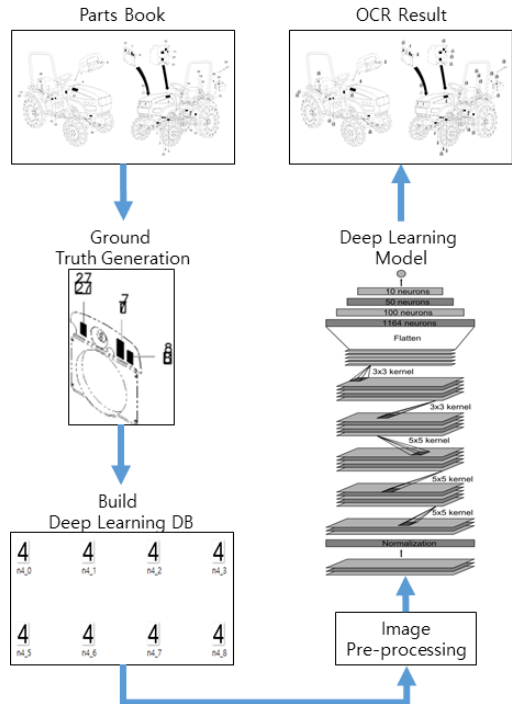


그림 2. 제안하는 OCR 구조

Fig. 2 The proposed OCR structure

본 논문에서는 침식 연산을 먼저 수행한 후에 팽창 연산을 수행하는 열림 연산을 사용했다. 객체를 축소하는 최소값 필터 침식 연산을 통해 지시선과 도형을 제거하고, 팽창 연산을 수행하여 인식해야 할 문자 및

숫자가 침식 연산 과정에서 축소된 것을 팽창한다. 그림 3은 전처리 과정에서 일반적으로 사용되는 형태소 마스크이다. 본 논문에서는 3x3의 십자(cross)형태의 마스크를 사용하였다.

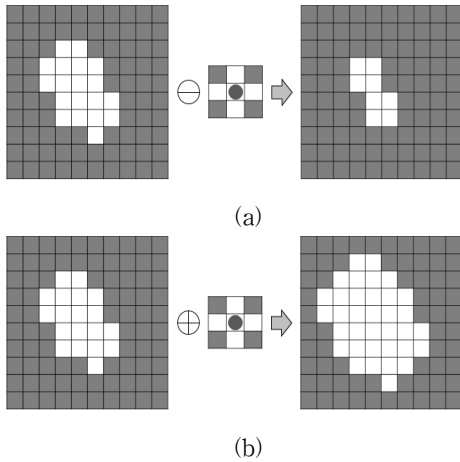


그림 3. 수학적 형태학 연산 처리 결과 (a) 침식연산, (b) 팽창연산
 Fig. 3 Results of mathematical morphology (a) Erosion, (b) Dilation

그림 4는 (a)는 서비스 부품도면의 원 이미지이며, (b)는 전처리 과정을 수행하지 않고 추론 한 결과이다. ©는 전처리를 수행한 결과이다. (b)에서 지시선 및 도형을 “1” 또는 “7”로 둥근 모양의 도형은 “0”으로 오검출 및 오인식 되지만, (c)에서는 대부분의 지시선과 도형이 제거된 것을 확인할 수 있다.

IV. 시뮬레이션 결과 및 검토

4.1 실험 환경

제안하는 OCR 성능 검증을 위하여 부품도면에서 0부터 9까지 28x28 크기의 숫자를 31,249개를 추출하여 데이터셋을 구성하였다. Ubuntu 16.04, NVIDIA Titan X 12GB 환경에서 축소 변형 VGG모델을 사용하여 학습한다. Windows 10, Visual studio .NET C# 환경에서 OpenCV 3.4 버전을 이용하여 전처리 과정을 수행하고 학습된 모델을 이용하여 추론하는 애플리케이션을 개발하였다.

4.2 실험 결과

성능 평가 사용한 부품도면의 해상도는 3,000x1,800 이다. 그림 5는 전처리와 결합한 축소 변형 VGG OCR를 적용한 결과이다.

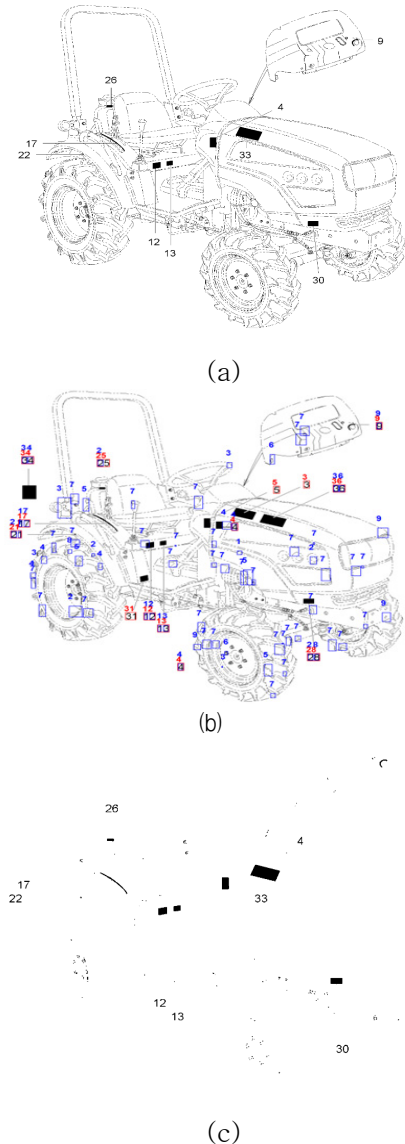


그림 4. 서비스 부품도면에 대한 OCR 및 전처리 결과. (a) 원본 영상, (b) OCR 처리 결과 (c) 수학적 형태학 필터링 적용 결과
 Fig. 4 Results of OCR and pre-processing. (a) original image, (b) a result of OCR (c) a result of mathematical morphology filtering

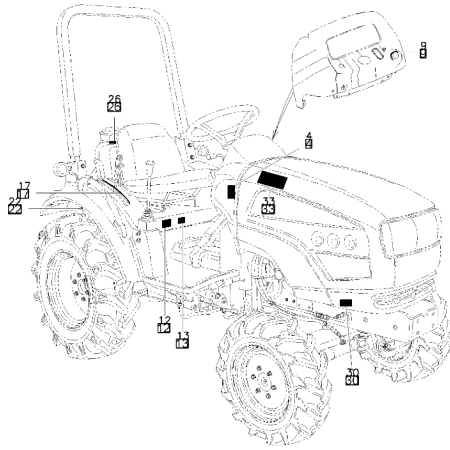


그림 5. 제안한 OCR 결과

Fig. 5 The Result of the proposed OCR

입력 도면 데이터에 대한 결과를 텍스트 파일로 로그 파일을 제작하여 기존에 있는 GT텍스트 파일과 비교하여 TP(True Positive), TN(True Negative), FP(False Positive), FN(False Negative)을 계수하여 식 (1), (2)와 같이 인식률(recall)과 정확도(precision)를 계산한다. 그리고, 인식률과 정확도의 조화 평균으로 F1 척도를 식 (3)과 같이 계산하여 비교한다.

$$R = \frac{TP}{TP + FP} \quad (1)$$

$$P = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2RP}{R + P} \quad (3)$$

표 2는 부품도면에 대한 OCR 성능을 비교한 결과이다. 서비스 부품도면 100장을 활용하여 OCR 성능을 비교한다. 실험 결과 데이터 입력 크기가 224×224인 기존의 VGG-16 모델을 사용했을 때 인식률과 정확도가 73.89%, 정확도 58.0% 이며, 축소 변형 VGG를 적용하면 인식률과 정확도가 각각 96.04%, 87.01%로 VGG-16 모델에 비하여 우수한 것을 확인할 수

있다. 그리고, 제안하는 OCR, 즉, 수학적 형태학 전처리와 결합한 축소 변형 VGG 모델의 경우 인식률과 정확도가 각각 95.57%, 92.82%로 축소 변형 VGG에 비하여 오검출 및 오인식이 현저히 개선되었다.

표 2. 부품도면 OCR 성능 비교
Table 2. Performance comparison of OCR of service parts book

Models	VGG-16	Mimi-VGG	proposed OCR
Recall	73.89%	96.04%	95.57%
Precision	58.00%	87.01%	92.82%
F1 score	66.10%	91.30%	94.17%

V. 결 론

본 논문은 기계 산업의 서비스 부품도면의 문자와 숫자 인식률을 개선을 위하여 수학적 형태학 전처리와 축소 변형 VGG 모델을 적용하는 방법을 제안하였다. 수학적 형태학 전처리 과정은 도면의 지시선과 부품이 오인식되어 인식률을 저하시키는 요인을 제거하며, 도면 숫자 크기를 고려해 기존 VGG-16이 아닌 축소 변형된 VGG 모델을 적용하여 정확도를 개선시킨다. 서비스 부품도면 100장에 대한 OCR 성능 평가를 F1 척도로 비교한 결과, 제안하는 OCR인 수학적 형태학 전처리와 결합한 축소 변형 VGG의 결과가 94.17%로 가장 성능이 우수함을 확인하고, 인식률(Recall) 95.57%, 정확도(Precision) 92.82%의 결과를 얻었다. 제안한 OCR이 서비스 부품도면 숫자 인식에 효과적으로 활용할 수 있음 확인하였다.

감사의 글

본 논문은 중소벤처기업부(S-BOM 기반 애프터마켓용 파츠북 서비스 ITEM 시스템 개발)와 2018년 BB21+ 사업에 의하여 지원되었음

References

- [1] J. Jo and H. Yang, "A Car License Plate Recognition Using Colors Information, Morphological Characteristic and Neural Network," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 25, no. 1, June 2010, pp. 304-308.
- [2] S. Park "An Effective Method of Product Number Detection from Thick Plates," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 25, no. 1, Jan. 2015, pp. 139-148.
- [3] S. Kim, S. Lee, S. Lee, and S. Lee "Household storage service through Optical Character Recognition(OCR)," *In Proc. Communications of the Korean Institute of Information Scientists and Engineers*, Pusan, Korea, Dec. 2017, pp. 377-379.
- [4] P. A. Wankhede, and S. W. Mohod, "A different image content-based retrievals using OCR techniques," *Proc. of Int. Conf. of Electronics, Communication and Aerospace Technology*, Coimbaore, India, Apr. 2017, pp. 20-22.
- [5] R. Smith, "An overview of the tesseract OCR engine," *Proc. of Int. Conf. on Document Analysis and Recognition*, Parana, Brazil, Sept. 2007, pp. 629-633.
- [6] X. Zhang, J. Zou, K. He, and J. Sun, "Accelerating very deep convolutional networks for classification and detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 38, 2015, pp. 1943-1955
- [7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. of Int. Conf. Learning Representations ICLR 2015*, San Diego, USA, May 2015.
- [8] X. C. Yin, X. Yin, K. Huang, and H. W. Hao, "Robust text detection in natural scene images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, 2014, pp. 970-983.
- [9] R. M. Haralick, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, July 1987, pp. 532-550,
- [10] L. Vincent, "Morphological area openings and closings for greyscale images," *Proc. NATO Shape in Picture Workshop*, Driebergen, The Netherlands, Sept. 1992, pp. 197-208.

저자 소개

신희란(Hee-Ran Shin)



2016년 동의대학교 멀티미디어 공학과 졸업(학사)
 2017년 ~ 2018년 웹시스템 프로 그래머
 2018년 6월 ~ 현재 경성대학교 산학협력단 연구원
 ※ 관심분야 : 컴퓨터비전, 머신러닝, 딥러닝

이상협(Sang-Hyeop Lee)



2016년 동의대학교 멀티미디어 공학과 졸업(학사)
 2018년 동의대학교 멀티미디어공 학석사 졸업(석사)
 2018년 6월 ~ 현재 경성대학교 전자공학과 박사 과정 재학 중
 ※ 관심분야 : 적응신호처리, 머신러닝, 컴퓨터비전

박장식(Jang-Sik Park)



1992년 부산대학교 전자공학과 졸업(학사)
 1994년 부산대학교 대학원 전자 공학과 졸업(석사)
 1999년 부산대학교 대학원 전자공학과 졸업(박사)
 1997년 ~ 2011년 동의과학대학교 전자학과 교수
 2011년 ~ 현재 경성대학교 전자공학과 교수
 2016년 ~ 현재 TTA PG-427 CCTV표준 의장
 ※ 관심분야 : 적응신호처리, 심층학습, 머신러닝, 컴퓨터비전, 임베디드 시스템

송중관(Jong-Kwan Song)



1989년 부산대학교 전자공학과 졸업(학사)
 1991년 한국과학기술원 대학원 전자공학과 졸업(석사)
 1999년 한국과학기술원 대학원 전자공학과 졸업 (박사)
 1997년 ~ 현재 경성대학교 전자공학과 교수
 ※ 관심분야 : 영상 처리 및 부호화, 심층학습, 머 신러닝, 컴퓨터비전