

일반논문 (Regular Paper)

방송공학회논문지 제27권 제3호, 2022년 5월 (JBE Vol.27, No.3, May 2022)

<https://doi.org/10.5909/JBE.2022.27.3.391>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

YOLOv4 기반의 소형 물체탐지기법을 이용한 건설도면 내 철강 자재 문자 검출 및 인식기법

심 지 우^{a)}, 우 회 조^{a)}, 김 윤 환^{b)}, 김 응 태^{a)*}

Character Detection and Recognition of Steel Materials in Construction Drawings using YOLOv4-based Small Object Detection Techniques

Ji-Woo Sim^{a)}, Hee-Jo Woo^{a)}, Yoonhwan Kim^{b)}, and Eung-Tae Kim^{a)*}

요 약

최근 딥러닝 기반의 객체 검출 및 인식 연구가 발전해가면서 산업 및 실생활에 적용되는 범위가 넓어지고 있다. 건설 분야에도 딥러닝 기반의 시스템이 도입되고 있지만 아직은 미온적이다. 건설 도면에서 자재 산출이 수작업으로 이뤄지고 있어 많은 소요시간과 부정확한 계산 결과로 잘못된 물량산출의 거래가 생길 수 있다. 이를 해결하기 위해서 빠르고 정확한 자동 도면 인식시스템이 필요하다. 따라서 본 논문은 건설도면 내 철강 자재를 검출하고 인식하는 인공지능기반 자동 도면 인식 시스템을 제안한다. 빠른 속도의 YOLOv4 기반에 소형 객체 검출성능을 향상하기 위한 복제 방식의 데이터 증강기법과 공간집중 모듈을 적용하였다. 검출한 철강 자재 영역을 문자 인식한 결과를 토대로 철강 자재를 적산한다. 실험 결과 제안한 방식은 기존 YOLOv4 대비 정확도와 정밀도를 각각 1.8%, 16% 증가시켰다. 제안된 방식의 Precision은 0.938, Recall은 1, $AP_{0.5}$ 는 99.4%, $AP_{0.5:0.95}$ 는 68.8%의 향상된 결과를 얻었다. 문자 인식은 기존 데이터를 사용한 인식을 75.6%에 비해 건설도면에 사용되는 폰트에 맞는 데이터 세트를 구성하여 학습한 결과 99.9%의 인식률을 얻었다. 한 이미지 당 평균 소요시간은 검출 단계는 0.013초, 문자 인식은 0.65초, 적산 단계는 0.16초로 총 0.84초의 결과를 얻었다.

Abstract

As deep learning-based object detection and recognition research have been developed recently, the scope of application to industry and real life is expanding. But deep learning-based systems in the construction system are still much less studied. Calculating materials in the construction system is still manual, so it is a reality that transactions of wrong volume calculation are generated due to a lot of time required and difficulty in accurate accumulation. A fast and accurate automatic drawing recognition system is required to solve this problem. Therefore, we propose an AI-based automatic drawing recognition accumulation system that detects and recognizes steel materials in construction drawings. To accurately detect steel materials in construction drawings, we propose data augmentation techniques and spatial attention modules for improving small object detection performance based on YOLOv4. The detected steel material area is recognized by text, and the number of steel materials is integrated based on the predicted characters. Experimental results show that the proposed method increases the accuracy and precision by 1.8% and 16%, respectively, compared with the conventional YOLOv4. As for the proposed method, Precision performance was 0.938. The recall was 1. Average Precision $AP_{0.5}$ was 99.4% and $AP_{0.5:0.95}$ was 67%. Accuracy for character recognition obtained 99.9% by configuring and learning a suitable dataset that contains fonts used in construction drawings compared to the 75.6% using the existing dataset. The average time required per image was 0.013 seconds in the detection, 0.65 seconds in character recognition, and 0.16 seconds in the accumulation, resulting in 0.84 seconds.

Keyword : Construction Drawings, Text Recognition, YOLO, Data Augmentation, Spatial Attention

Copyright © 2022 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

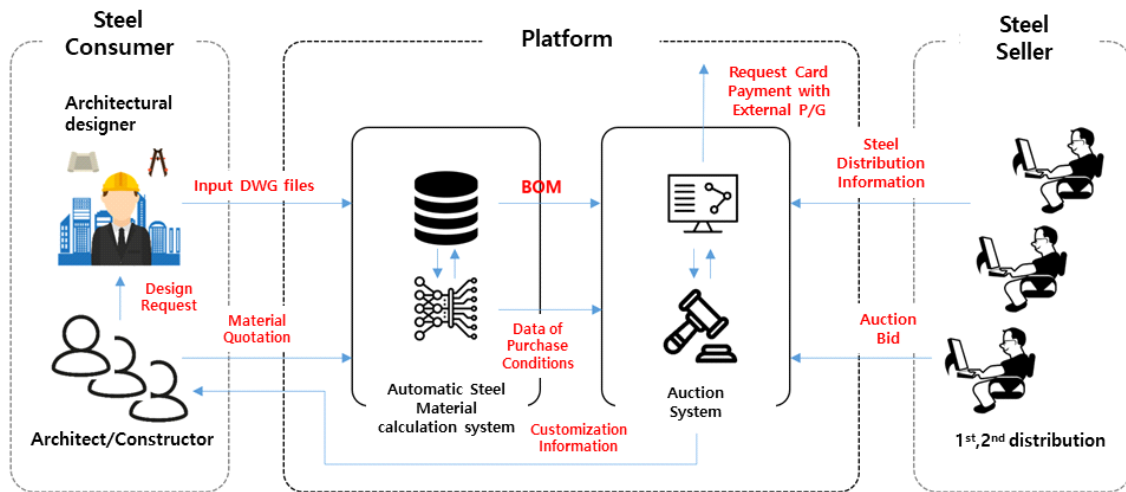


그림 1. 철강 재료의 자동 적산 시스템 기반 경매시스템

Fig. 1. Auction system based on automatic quantity surveying of steel materials

1. 서론

딥러닝에 관한 연구는 나날이 발전해가고 있으며 그중에서도 객체 검출과 인식 관련 연구가 활발히 진행되고 있어^[1-8], 이미 산업과 실생활의 많은 분야에 적용되고 있다. 객체 검출 기술이 사람의 일손을 대신하며 일의 효율성에 도움을 주고 있는 것은 사실이지만, 아직은 사람이 직접 개입하는 일보다는 정확도가 떨어지는 문제점을 가지고 있다. 그런데도 사람을 대체하여 일할 수 있고 같은 효율을 내는 인공지능을 구현하기 위해 연구는 계속되고 있다.

AI(Artificial Intelligence)가 활발히 연구되고 산업 및 실생활에 적용되는 중심에는 4차 산업혁명이 있다. 4차 산업혁명은 모든 산업에서 딥러닝 등 기계학습과 빅데이터에

기반한 인공지능을 활용하여 기술 및 산업구조가 초지능화 되는 것을 목표로 하고 있다. 하지만 현재 국내 건설 분야는 AI 및 빅데이터를 포함한 4차 산업혁명 관련 기술의 도입에 미온적이며 국내 건설특성을 반영한 BIM(Building Information Modeling) S/W의 개발이 필요하다. 기존 건설에 필요한 원자재 시장은 전형적인 인적기반의 재래식 유통구조로 거래처 간 제품 및 가격에 대한 객관적인 신뢰가 불투명하다. 그중에서도 건설의 초입 단계에서 건설자재를 산출하는 과정은 아직 수작업으로 진행되고 있다. 이는 많은 소요시간과 정확한 적산¹⁾의 어려움이 있어 잘못된 물량산출의 거래가 생길 수 있다. 따라서 비효율적인 재작업을 줄이기 위해서는 그림 1에서 보는 바와 같이 건설도면을 구성하는 건설자재의 정보를 자동으로 인식하는 인공지능 기반 적산시스템이 필요하다. 이 시스템은 건축 및 설계 사무소가 건축주나 시공사로부터 설계 의뢰를 받은 설계 캐드 파일을 자동 자재산출시스템을 통해 BOM(Bill of Material)을 제공하고 1, 2차 철강 공급자가 제공하는 최적 가격을 신속하고 정확하게 건축주 및 시공사에 제공할 수 있는 시스템이다.

최근 딥러닝 기반의 도면 인식에 관한 연구가 진행되고 있는데, 해당 연구는 이미지 형식의 P&ID 도면을 지능형 P&ID 도면으로 변환을 위한 심볼 인식에 관련한 것에 그친

a) 한국공학대학교 전자공학부(Department of Electronics Engineering, Tech University of KOREA)

b) ㈜우성철강산업(WooSung Steel Inc.)

‡ Corresponding Author : 김응태(Eung-Tae Kim)

E-mail: etkim@tukorea.ac.kr

Tel: +82-31-8041-0488

ORCID: <http://orcid.org/0000-0001-5984-0045>

※ This work was supported by the Technology development Program (S3025098) funded by the Ministry of SMEs and Startups(MSS, Korea).
· Manuscript April 7, 2022; Revised May 10, 2022; Accepted May 10, 2022.

1) 공사에 소요되는 재료, 노무의 수량 · 단가, 품의 수량 등을 계산하는 것.

다^[9]. 이와 비슷한 도면을 인식하는 연구들은 있지만, 딥러닝 기반으로 도면에서 건설자재를 검출하여 자동 인식하는 연구는 찾아볼 수 없다. 따라서 본 논문에서는 건설도면 내에 명시된 철강 자재를 검출하고 인식하는 자동 도면 인식 적산 시스템 구현한다. 이때 검출하고자 하는 철강 자재 영역이 도면 이미지의 크기에 비교하여 매우 작은 영역을 차지하고 있다. 소형 객체의 검출 정확도는 아직 객체 검출 연구에서 해결해야 할 난제로 남겨져 있다. 따라서 소형 객체의 검출 정확도를 증가시키기 위해 데이터 증강기법과 공간집중모듈을 철강 자재 검출에 적합하게 변형시켜 적용하여 개선하고자 한다.

II. 제안하는 건설도면 인식시스템

본 논문에서 제안하는 자동 건설도면 인식시스템의 전체적인 과정은 그림 2와 같다. 도면 이미지를 입력하여 이미지 내의 철강 자재를 검출하고 검출할 영역들을 문자 인식을 용이하게 만들기 위해서 이미지 전처리를 거쳐 비교적 정확한 결과를 내는 문자 인식을 통해 철강 자재 명을 추출 후 적산하여 엑셀 파일에 정리한다. 큰 분류로 철강 자재 검출과 인식으로 나누어 소개하고자 한다.

1. 도면 내의 건설자재 검출

본 논문에서 제안하는 시스템은 산업현장에 사용되는 것을 고려하여 1개의 GPU만을 사용해도 학습속도가 빠르고 정확도가 보장되는 YOLOv4 기법^[8]을 사용하였다. YOLO 모델의 장점은 기존 2-stage 방식인 R-CNN^[2], Fast R-CNN^[3], Faster R-CNN^[4] 네트워크 대비 빠른 검출 속도이다. 그렇지만 단점으로 소형 객체에 대한 검출성능이 다른 모델들과 비교하여 떨어진다는 한계를 갖는다. 따라서 1584x1224 이상의 픽셀 크기를 가지는 도면 이미지보다 약 0.013배의 크기를 갖는 소형의 문자와 숫자가 결합하는 텍스트 형태의 철강 자재 문자 영역을 검출해야 한다. 이는 평균 19x12 픽셀크기이내의 소형 객체에 해당하며, 이런 객체 영역의 검출성능을 높이는 것이 필수이다. 따라서 기존의 YOLOv4에 소형 객체 검출성능을 향상하는 데이터 증강 기법^[10-11]과 Spatial Attention^[12-13]의 적용을 제안하며 전체적인 학습 네트워크는 그림 3과 같다.

소형 객체 검출 정확도가 낮은 이유에는 데이터 부족에 대한 문제가 있다. 도면 이미지에 기재된 철강 자재에 대한 데이터 세트는 존재하지 않기 때문에 직접 구성해야 한다. 이때 모든 경우의 철강 자재에 대한 데이터 세트를 구성하

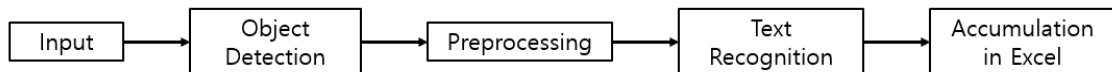


그림 2. 제안하는 건설도면 인식시스템
Fig. 2. Proposed construction drawing recognition system

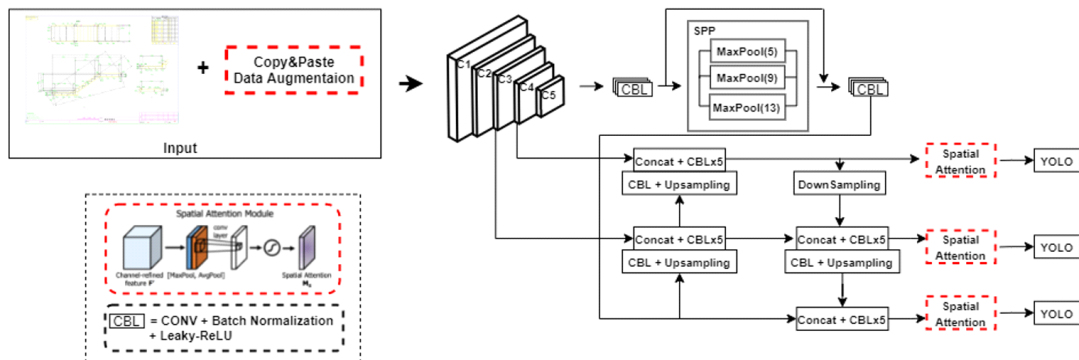


그림 3. 제안된 학습 네트워크
Fig. 3. Proposed training network

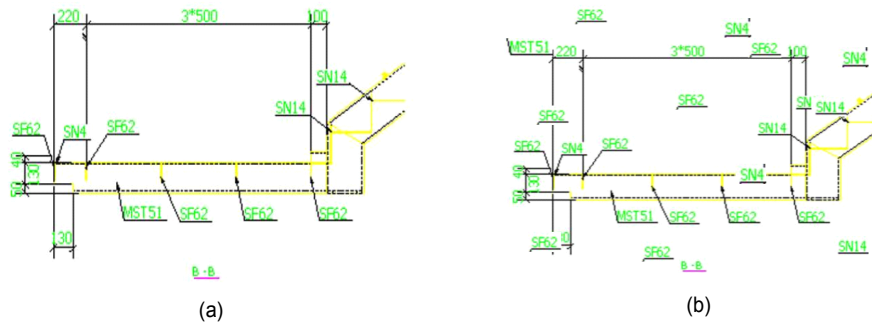


그림 4. 복제 방식의 데이터 증강기법
Fig. 4. Copy & paste data augmentation

기에는 한계가 있다. 따라서 자체적으로 철강 자재 데이터를 증가시키고 다양한 경우를 연출하기 위해 복제 방식의 데이터 증강기법을 적용한다. 그림 4는 전체 도면의 일부를 확대하여 표기한 것으로 그림 4의 (a)에서 철강 자재에 해당하는 영역을 복사해서 그림 4의 (b)와 같이 무작위한 위치에 복사하는데 이때 전체 이미지 영역을 벗어나지 않게 붙여 넣어 오류를 피한다. 객체를 복사해서 붙여 넣을 때, 도면에서의 다양한 경우의 학습을 위해 객체들끼리 겹치거나 다른 정보들을 가리거나 하는 다양한 형태가 나타나는 것을 허용하며 객체를 복제하는 횟수를 증가시킬 수 있다.

또한, 소형 객체의 정확도가 낮은 이유는 작은 영역에서 추출할 수 있는 특징이 제한적이기 때문이다. 따라서 추출되는 특징들에서 중요한 특징에 더 집중해야 할 필요가 있어 그림 3의 오른쪽 빨간 실선과 같이 검출 단계 이전에 Spatial Attention 적용을 제안한다. Spatial Attention은 이

전에 언급한 것과 같이 정보가 담긴 부분이 어디에 있는지 집중한다. 기존의 Spatial Attention은 Average Pooling과 Max Pooling으로 연산된 값에 7x7의 컨볼루션 연산을 적용하였지만 본 논문에서는 큰 이미지에 비해 철강 자재의 영역은 최소 0.013배 작기 때문에 매우 작은 영역에서마다 정보를 추출하기 위해 3x3의 컨볼루션으로 대체하며 식 (1)과 같이 표기한다. 식 (1)의 M_s 는 이전의 연산값, σ 는 sigmoid function, F_{avg}^s 는 Average Pooling을 적용한 값, F_{max}^s 는 Max Pooling을 적용한 값을 의미한다.

$$M_s(F) = \sigma(f^{3 \times 3}([AvgPool(F); MaxPool(F)])) \quad (1)$$

$$= \sigma(f^{3 \times 3}([F_{avg}^s; F_{max}^s]))$$

이외에도 소형 객체에 대한 정확도를 증가시키기 위해 앵커 박스의 크기를 수정하였다. 기존 YOLOv4에서 사용

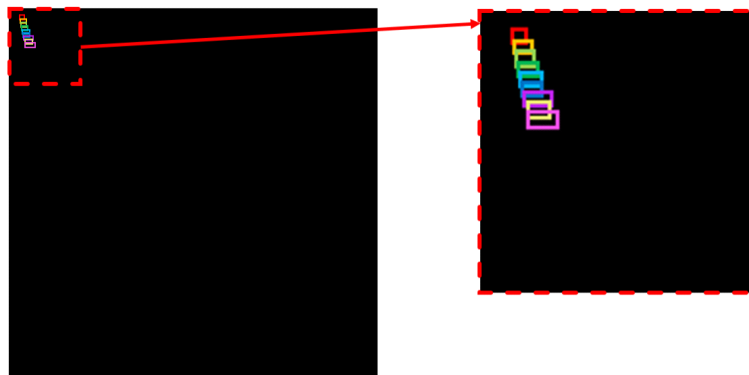


그림 5. 철강 자재 텍스처 검출용 앵커 박스 모양
Fig. 5. Anchor box style for detection of steel material texture

하는 MSCOCO 데이터 세트를 기반으로 사용하는 앵커 박스의 크기는 12, 16, 19, 36, 40, 28, 36, 75, 76, 55, 72, 146, 142, 110, 192, 243, 459, 401이다. 하지만 이는 도면 이미지 내의 철강 자재 텍스트를 검출하기에는 적합한 형태가 아니다. 따라서 그림 5와 같이 K-평균 클러스터링을 철강 자재 데이터 세트에 적용하여 12, 12, 15, 11, 15, 13, 17, 11, 19, 11, 18, 13, 23, 11, 20, 13, 25, 13로 수정하였다.

2. 철강 자재 문자 인식

문자 인식에는 딥러닝 기반의 OCR-net인 TRBA(TPS-ResNet-BiLSTM-Attn)^[15] 네트워크를 사용한다. TRBA 네트워크는 학습된 가중치를 제공한다. 하지만 건설도면에 사용하는 폰트를 적용하는 문자를 제공되는 가중치를 적용하면 낮은 정확도가 나오기 때문에 도면 내의 자재 명으로 데이터 세트를 구성하여 학습하였다. 검출된 철강 자재 영역은 이전에 언급한 것과 같이 매우 작은 크기를 가지기 때문에 많은 특징을 추출하기 위해 bicubic 기법으로 2배

확대하며 문자 인식 성능을 높이기 위해 전처리 과정을 거친다. 검출된 철강 자재 영역에는 도면 설계에 사용된 검정선과 같은 문자 인식에 불필요한 정보가 포함되어 있으며 그림 6의 (a)에 해당한다. 이것을 제거하여 그림 6의 (b)와 같이 변형시켜 입력 이미지로 사용한다.



그림 6. 문자 인식에 사용하는 입력 이미지
Fig. 6. Input image used for character recognition

III. 데이터 세트 구성

1. 제안한 SMD 데이터 세트

객체에 대한 높은 검출성능을 갖기 위해서는 적합한 데이터 세트를 적용하여 학습하는 것이 중요하다. 현재 객체

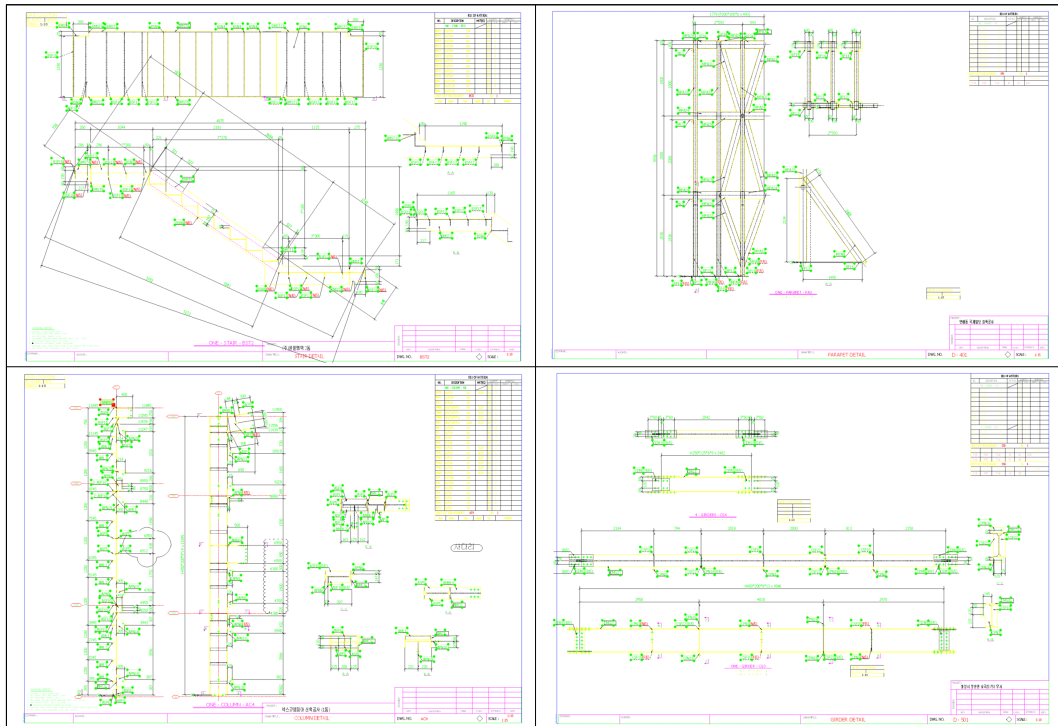


그림 7. SMD 데이터 세트의 철강 자재를 라벨링 한 도면 이미지 예
Fig. 7. Example of a drawing image labeled steel materials in an SMD dataset

검출에 사용되는 데이터 세트로는 MSCOCO와 VOC가 대중적이며 소형 객체 검출에 사용되는 데이터 세트로는 KITTI, VisDrone과 Udacity가 있지만 도면 이미지에 적용하여 철강 자재를 검출하기에는 적합하지 않다. 따라서 건설 도면에서의 철강 자재 검출을 목적으로 본 논문에서는 실제 건설 현장에서 사용된 건설 도면 데이터들을 사용하여 철강 자재 영역을 라벨링하여 SMD(Steel Materials in Drawings) 데이터 세트를 직접 만들어 학습하였다. SMD 데이터 세트에는 다양한 건물의 도면에 철강 자재들이 포함된다. SMD의 세부 구성은 표 1과 같으며 예시는 그림 7과 같다.

표 1. SMD 데이터 세트
Table 1. SMD Dataset

Category	Train	Validation	Test
Images	719	26	80
Total	825		

2. N_SMD 데이터 세트

건설도면에서 사용하는 폰트를 적용하는 문자를 인식하기 위한 목적으로 N_SMD(Names of Steel Materials in Drawings) 데이터 세트를 구성하였다. 검출한 철강 자재 영

표 2. N_SMD 데이터 세트
Table 2. N_SMD Dataset

Category	Train	Validation	Test
Images	10000	3000	4582
Total	17582		

SF22	AP2	EM25	BSF25
AMBK2	ASF1	AMC1	APN1
ABP2	CP3	BSF28	EM1
PN22	PN10	SF2	MC16
PN9	BSF25	BMBK20	ASF24
SN6	SN25	BSF25	ASF6

그림 8. N-SMD 데이터 세트의 예
Fig. 8. Example of N-SMD dataset

역에 bicubic 기법을 적용하여 2배 확대 후 자재 명과 관계 없는 정보에 해당하는 픽셀을 제거한 이미지를 데이터 세트에 포함하며 그림 8과 같다.

IV. 모의실험 결과

1. 철강 자재 검출 실험

본 논문에서 제안하는 철강 자재 검출 및 인식을 통한 자동 적산 시스템에서 검출 부분과 인식 부분을 나누어 모의실험한 후 전체 시스템에 대해서 실험하였다. Linux 기반의 Ubuntu 18.04 LTS Server/ Intel Xeon Gold 640R CPU @ 2.40GHz/ GeForce RTX 3090 24GB 환경에서 실험하였다.

본 모의실험에서는 제안하는 SMD 데이터 세트만을 기반으로 성능을 분석하였으며 철강 자재에 해당하는 하나의 객체만 검출하므로 AP값으로 평가한다. 먼저 기존 YOLOv4와 동일한 하이퍼 파라미터를 사용하였으며 복제 방식의 데이터 증강기법을 1~3번까지 복제하는 횟수를 증가시켜 적용한 결과를 비교하였다. 본 논문에서 검출성능을 평가할 때 Precision(정밀도), Recall(재현율), AP(Average Precision)의 지표들을 사용한다. Precision이란 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율이다. Recall이란 실제 True인 것 중에서 모델이 True라고 예측한 것의 비율이다. 검출되어야 할 객체 중에서 제대로 검출된 것의 비율을 뜻한다. AP은 Recall을 0부터 0.1 단위로 증가시켜서 1까지 증가시킬 때 필연적으로 Precision이 감소하는데 단위마다 Precision 값을 계산하여 평균을 내어 계산한다. 하나의 클래스마다 AP값을 계산할 수 있다. $AP_{0.5}$, $AP_{0.5:0.95}$ 를 사용해서 성능을 평가하는데 여기서 0.5와 0.95는 IoU (Intersection over Union)의 임계값의 범위를 나타낸다. IoU란 예측된 bounding box와 실제 bounding box의 겹치는 구간을 그 두 개 구간의 합으로 나눈 것이다.

표 3에서는 복제 방식의 데이터증강기법을 적용한 결과이다. 기존 YOLOv4의 결과, Precision은 0.779, Recall은 0.998, $AP_{0.5}$ 는 99.4%, $AP_{0.5:0.95}$ 67%가 나왔다. 이에 비교하여 복제 방식의 데이터 증강기법을 1번 적용했을 때는

Precision은 0.816, Recall은 1, $AP_{0.5}$ 는 99.5%, $AP_{0.5:0.95}$ 67.6%의 결과가 나왔고, 2번 적용했을 때는 Precision은 0.822, Recall은 1, $AP_{0.5}$ 는 99.5%, $AP_{0.5:0.95}$ 68%의 결과가 나왔고, 3번 적용했을 때는 Precision은 0.842, Recall은 1, $AP_{0.5}$ 는 99.4%, $AP_{0.5:0.95}$ 68%의 결과가 나왔다. 결과적으로 기존 YOLOv4만을 사용하는 것보다 복제 방식의 데이터 증강기법을 적용하는 것이 성능을 높일 수 있으며 복제의 횟수를 증가할수록 성능이 개선됨을 확인할 수 있었다.

표 3. 복제 방식의 데이터 증강기법 적용 결과
Table 3. Result of applying the Copy&Paste Data augmentation

	Precision	Recall	$AP_{0.5}$	$AP_{0.5:0.95}$
YOLOv4	0.779	0.998	99.4	67.0
Aug x1	0.816	1	99.5	67.6
Aug x2	0.822	1	99.5	68
Aug x3	0.842	1	99.4	68

다음은 표 4에서 보는 바와 같이 Spatial Attention 모듈을 적용한 성능을 비교하였다. 기존 Spatial Attention 모듈의 7x7의 컨볼루션 연산을 적용했을 때는 Precision은 0.834, Recall은 0.998, $AP_{0.5}$ 는 99.5%, $AP_{0.5:0.95}$ 66.1%의 결과가 나왔고, 철강 자재 검출에 적합하게 컨볼루션 연산의 크기를 3x3으로 줄여 적용했을 때는 Precision은 0.948, Recall은 1, $AP_{0.5}$ 는 99.5%, $AP_{0.5:0.95}$ 67.1%의 결과가 나왔다. 결과적으로 기존 YOLOv4와 비교했을 때 Spatial Attention 모듈을 적용함으로써 성능을 개선할 수 있으며 3x3의 컨볼루션 연산을 적용하는 것이 더 크게 성능 개선된 것을 확인하며 소형의 철강 자재 영역 검출에 적합한 개선임을 확인하였다.

표 4. Spatial Attention 적용 결과
Table 4. Result of applying Spatial Attention

	Precision	Recall	$AP_{0.5}$	$AP_{0.5:0.95}$
YOLOv4	0.779	0.998	99.4	67.0
SA_3x3	0.948	1	99.5	67.1
SA_7x7	0.834	0.998	99.5	66.1

마지막은 표 5에서 보는 바와 같이 복제 방식의 데이터 증강기법과 Spatial Attention 모듈을 적용한 성능을 비교하였다. 위의 결과에 따라 3x3의 컨볼루션 연산을 하는

Spatial Attention 모듈을 고정으로 적용하여 복제 방식의 데이터 증강기법의 적용 횟수를 증가시켜 결과를 비교하였다. 1번 적용했을 때는 Precision은 0.931, Recall은 1, $AP_{0.5}$ 는 99.5%, $AP_{0.5:0.95}$ 67.4%의 결과가 나왔고, 2번 적용했을 때는 Precision은 0.945, Recall은 1, $AP_{0.5}$ 는 99.5%, $AP_{0.5:0.95}$ 68%의 결과가 나왔고, 3번 적용했을 때는 Precision은 0.938, Recall은 1, $AP_{0.5}$ 는 99.4%, $AP_{0.5:0.95}$ 68.8%의 결과가 나왔다. 결과적으로 기존 YOLOv4에 비교해서 복제 방식의 데이터 증강기법과 Spatial Attention 모듈을 적용했을 때 성능이 개선됨을 알 수 있으며, 복제 횟수를 3번 적용했을 때 Precision, Recall, $AP_{0.5}$, $AP_{0.5:0.95}$ 모두가 가장 좋은 성능을 내는 것을 확인할 수 있었다. 따라서 제안하는 시스템에 3회 복제 방식의 데이터 증강기법과 Spatial Attention 모듈을 적용한다.

표 5. 복제 방식의 데이터 증강기법과 Spatial Attention 적용 결과
Table 5. Result of applying the Copy&Paste Data augmentation and Spatial Attention

	Precision	Recall	$AP_{0.5}$	$AP_{0.5:0.95}$
YOLOv4	0.779	0.998	99.4	67.0
SA7x7 + Aug	0.816	0.999	99.5	67.1
SA + Aug x1	0.931	1	99.5	67.4
SA + Aug x2	0.945	1	99.5	68
SA + Aug x3	0.948	1	99.5	68.8

그림 9는 제안된 기법을 실제 도면에서 검출한 결과이다. 기존의 YOLOv4 모델만 사용하여 검출했을 경우 그림 9(a)에 나타난 것과 같이 표기하지 못한 것과 검출은 하였지만, 영역 전체를 다 검출하지 못한 것을 발견할 수 있었다. 또한, 학습 데이터 대부분이 철강 자재 영역이 초록색이기 때문에 노란색의 철강 자재명은 검출하지 못한 결과와 철강 자재가 아닌 다른 정보들과 겹쳐있는 경우 검출하지 못한 결과를 발견할 수 있었다. YOLOv4 모델에 복제 방식의 데이터 증강기법을 적용함으로써 기대한 학습시킨 일반적인 도면 데이터 외의 다양한 경우를 검출하는 것을 확인할 수 있었다. 그림 9(b)과 같이 노란색 철강 자재 영역을 검출하고 철강 자재가 아닌 다른 정보와 겹친 경우에서 철강 자재 영역을 검출하는 것을 확인할 수 있다. 또한, YOLOv4 모델에 7x7을 3x3 컨볼루션 연산으로 줄인 Spatial Attention 모

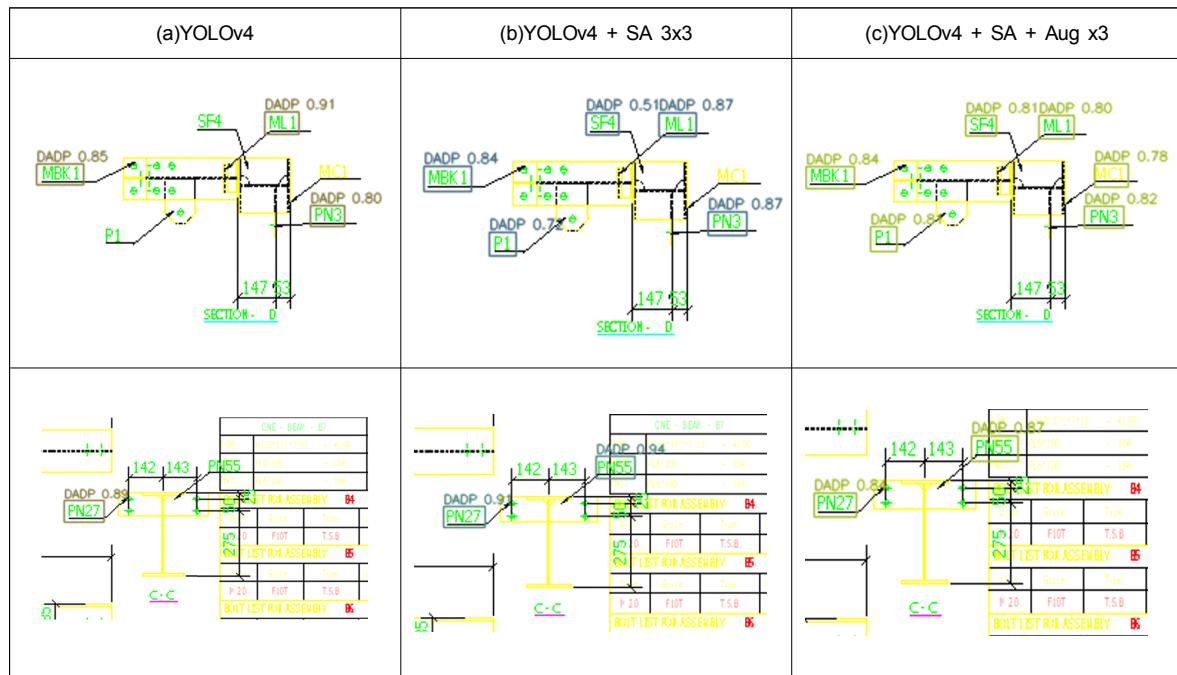


그림 9. 기존 YOLOv4와 제안하는 방식의 결과 비교

Fig. 9. Comparing the results of the proposed method with YOLOv4

들을 적용함으로써 기대한 소형 객체 검출의 성능 향상을 확인할 수 있었다.

가장 좋은 성능 결과가 나왔던 Spatial Attention과 복제 방식의 데이터 증강기법을 3번 적용했을 경우 그림 9(c)와 같다. 그림 9(b)와 같은 위치의 결과로, 검출하지 못했던 철강 자재 영역을 검출하였으며 노란색의 철강 자재 영역, 다른 정보와 겹친 철강 자재 영역 모두 검출한 결과를 확인할

수 있다.

2. 문자 인식 결과

문자 인식 부분은 TRBA 네트워크에 N_SMD 데이터 세트를 학습하였다. 검출된 영역을 입력하기 전 인식률을 높이기 위해 불필요한 정보를 삭제하는 전처리 과정을 포함

COCO-Text	N_SMD
ASF19 → ASF19	ASF19 → ASF19
EM25 → EM25	EM25 → EM25
EM1 → EM1	EM1 → EM1
PN9 → PN9	PN9 → PN9
AP2 → AP2	AP2 → AP2

그림 10. COCO-Text 데이터 세트와 N_SMD 데이터 세트 적용 결과

Fig. 10. Comparing the results of applying COCO-Text and N_SMD

하여 실험하였다. 기존의 COCO-Text 데이터 세트를 적용했을 때는 75.6%, N_SMD 데이터 세트를 적용했을 때는 99.9%의 정확도를 확인하였으며 표 6과 같다. 그림 10과 같이 COCO-Text 데이터 세트를 적용했을 때는 ‘1’을 ‘I’로, ‘5’를 ‘S’로, ‘9’를 ‘g’로, ‘2’를 ‘S’로 잘못 인식하는 오류가 발견되었다. 하지만 N_SMD 데이터 세트 적용으로 모든 오류의 경우가 개선됨을 확인할 수 있었다.

표 6. 문자 인식 결과
Table 6. Result of Recognition

	COCO-Text	N_SMD
Accuracy	75.6%	99.9%

3. 전체 시스템의 처리 속도 결과

제안한 시스템은 크게 세 가지 단계로 구분할 수 있다. 첫 번째는 건설도면 이미지에서 철강 자재 영역을 검출하는 것이다. 도면 이미지를 입력해서 검출된 영역의 좌표를 저장하는 단계까지를 의미한다. 이때 소요시간은 평균적으로 한 장당 0.033초가 소요되는 것을 확인하였다. 두 번째는 검출된 철강 자재 영역을 문자 인식하는 단계이다. 이 단계는 검출된 영역의 이미지를 저장 후 입력하여 문자 인식하여 예측된 문자들 도출하는 것으로 한 장에 담긴 영역의 수에 따라 소요시간이 다르지만, 평균적으로 한 장당 0.65초 소요되는 것을 확인할 수 있었다. 마지막 단계는 예측된 문자들을 기반으로 각 철강 자재들의 수량을 적산하여 엑셀 파일에 정리하여 저장하는 것까지를 의미하며 평균적으로 한 장당 0.016초 걸리는 것을 확인하였다. 각 단계 별 소요시간은 표 7과 같다. 전체적으로 제안된 시스템은 한 장당 철강 자재 검출, 인식, 적산처리까지 0.84초를 소요하므로 실시간 처리가 가능하다. 이는 사람이 수동적

표 7. 제안한 시스템의 처리 속도
Table 7. Processing Time of Proposed System

	Processing Time
Object Detection	0.033s
Text recognition	0.65s
Accumulation	0.16s
Total	0.84s

으로 처리할 수 있는 수십분에 비해 매우 빠르고 정확하게 처리할 수 있음을 확인할 수 있었다.

V. 결 론

건설도면에서의 자재 산출이 아직 수작업으로 이뤄지고 있어 많은 소요 시간과 부정확한 산출로 인한 눈속임 거러에 대한 문제를 해결하고자 본 논문에서는 자동 건설도면 인식시스템을 구현한다. 건설도면 내의 철강 자재 영역을 검출하여 문자 인식해서 철강 자재 명으로 적산하는 것이다. 산업 시스템에 구동되어야 하기 때문에 빠른 속도를 낼 수 있는 모델인 YOLOv4를 사용하였다. 하지만 기존 YOLOv4를 그대로 사용할 경우 작은 소형 객체 검출성능이 떨어지는 문제가 생긴다. 특히 건설도면 이미지에 비해 철강 자재 영역은 매우 작기때문에 소형 객체 검출성능 개선은 필수이다. 따라서 복제 방식의 데이터 증강기법과 3x3의 컨볼루션을 적용하는 Spatial Attention 모듈을 적용하고 철강 자재 영역에 맞게 앵커 박스의 크기를 조절하였다. 결과적으로 기존 YOLOv4만을 적용했을 때 Precision은 0.779, Recall은 0.998, $AP_{0.5}$ 는 99.4%, $AP_{0.5:0.95}$ 67%의 성능이 나온 반면, 3x3의 컨볼루션을 적용하는 Spatial Attention과 복제 방식의 데이터 증강기법을 3번 적용했을 때 Precision은 0.938, Recall은 1, $AP_{0.5}$ 는 99.4%, $AP_{0.5:0.95}$ 68.8%의 성능이 나왔다. 실험 결과 제안한 방식은 기존 YOLOv4 대비 정확도와 정밀도를 각각 1.8%, 16% 증가시켰다. 실제 이미지에서 결과를 비교했을 때 기존 YOLOv4만을 사용했을 때 검출하지 못했던 철강 자재 영역을 검출하고 학습된 데이터 이외의 다양한 상황들을 검출하였다. 이를 통해 소형 객체 검출성능을 개선하여 철강 자재 검출률을 높일 수 있었다. 검출한 철강 자재 영역을 TRBA 네트워크에 입력하여 본 논문에서 구성한 N_SMD 데이터 세트를 입력하여 얻은 가중치를 통해 철강 자재 명을 인식하였다. 결과적으로 99.9%의 인식률 성능을 보였다.

본 논문에서 제안하는 방식을 통해 건설도면 이미지 내의 철강 자재의 높은 검출률과 인식률을 기반으로 자동 건설자재 적산 시스템을 구축할 수 있다. 또한 고해상도 이미지에서 소형의 객체를 검출해야하는 다양한 비전시스템에

적용할 수 있으며 데이터 세트를 다량으로 구축할 수 없는 상황에서 복제방식의 데이터 증강기법을 적용하여 데이터를 증가시켜 학습함으로써 성능을 개선할 수 있다.

참 고 문 헌 (References)

- [1] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 580-587, 2014.
doi: <https://doi.org/10.1109/CVPR.2014.81>
- [2] Ross Girshick. "Fast R-CNN", In IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1440-1448, 2015.
doi: <https://doi.org/10.1109/ICCV.2015.169>
- [3] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", In Advances in Neural Information Processing Systems 28, 2015.
doi: <https://doi.org/10.1109/TPAMI.2016.2577031>
- [4] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.
doi: <https://doi.org/10.1109/CVPR.2016.91>
- [5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg. "SSD: Single Shot MultiBox Detector", In European Conference on Computer Vision, pp 21-37, 2016.
doi: https://doi.org/10.1007/978-3-319-46448-0_2
- [6] Jan Hosang, Rodrigo Benenson, Bernt Schiele. "Learning non-maximum suppression", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4507-4515, 2017.
doi: <https://doi.org/10.1109/CVPR.2017.685>
- [7] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection", arXiv preprint arXiv:2004.10934, 2020.
doi: <https://doi.org/10.48550/arXiv.2004.10934>
- [8] Chien-Yao Wang, Hong-Yuan Mark Liao, I-Hau Yeh, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh. "CSPNet: A New Backbone that can Enhance Learning Capability of CNN", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 390-391, 2020.
doi: <https://doi.org/10.1109/CVPRW50498.2020.00203>
- [9] Eun-seop Yu, Approach of object recognition from image format engineering drawings using deep learning, Department of Precision Mechanical Engineering Graduate School, Kyungpook National University Daegu, Korea, 2019.
- [10] Luis Perez, Jason Wang. "The Effectiveness of Data Augmentation in Image Classification using Deep Learning", Convolutional Neural Networks Vis. Recognit. 11:1-8. 2017.
doi: <https://doi.org/10.48550/arXiv.1712.04621>
- [11] Mate Kisantal, Zbigniew Wojna, Jakub Murawski, Jacek Naruniec, Kyunghyun Cho. "Augmentation for Small Object Detection", In arXiv preprint arXiv:1902.07296, 2019.
doi: <https://doi.org/10.5121/csit.2019.91713>
- [12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. "Attention Is All You Need", Advances in Neural Information Processing Systems, 30, 2017.
- [13] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon. "CBAM: Convolutional Block Attention Module", In Proceedings of European Conference on Computer Vision, pp.3-19, 2018.
doi: https://doi.org/10.1007/978-3-030-01234-2_1
- [14] Baoguang Shi, Xinggang Wang, Pengyuan Lyu, Cong Yao, Xiang Bai. "Robust Scene Text Recognition with Automatic Rectification", In IEEE Conference on Computer Vision and Pattern Recognition, pp. 4168-4176, 2016.
doi: <https://doi.org/10.1109/CVPR.2016.452>
- [15] Jeonghun Baek, Geewook Kim, Junyeop Lee, Sungrae Park, Dongyoon Han, Sangdoo Yun, Seong Joon Oh, Hwalsuk Lee, "What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis", In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4715-4723, 2019.
doi: <https://doi.org/10.1109/ICCV.2019.00481>

저 자 소 개



심 지 우

- 2019년 : 한국공학대학교 전자공학부 공학사
- 2022년 : 한국공학대학교 전자공학부 공학석사
- 2022년 3월 ~ 현재 : ㈜메디싱큐 기술연구소 연구원
- ORCID : <https://orcid.org/0000-0003-1964-3615>
- 주관심분야 : 딥러닝기반 영상처리, 컴퓨터 비전, 임베디드 시스템



우 희 조

- 2019년 : 한국공학대학교 전자공학부 공학사
- 2022년 : 한국공학대학교 전자공학부 공학석사
- 2022년 3월 ~ 현재 : 한화시스템 지능형 SW팀 연구원
- ORCID : <https://orcid.org/0000-0003-4956-6809>
- 주관심분야 : 딥러닝기반 영상처리, 경량화, 컴퓨터 비전, 임베디드 시스템



김 윤 환

- 2005년 : 홍익대학교 조선해양공학, 컴퓨터정보통신 공학사
- 2006년 3월 ~ 2013년 2월 : ㈜푸드락 대표이사
- 2013년 9월 ~ 2015년 11월 : ㈜CNO파트너스 본부장
- 2017년 1월 ~ 현재 : ㈜한국리서치엔연구원 책임연구원
- 2017년 3월 ~ 현재 : ㈜우성철강산업 대표이사
- 2020년 : 국민대학교 창업학 석사
- 2021년 : 전북대학교 융합기술경영학 박사과정
- ORCID : <https://orcid.org/0000-0001-8965-7823>
- 주관심분야 : AI기반 도면분석 알고리즘, 도면해석 자동적산 시스템



김 응 태

- 1991년 : 인하대학교 전자공학과 공학사
- 1993년 : KAIST 전기및전자공학과 공학석사
- 1999년 : KAIST 전기및전자공학과 공학박사
- 1998년 3월 ~ 2004년 2월 : ㈜LG전자 DTV연구소 책임연구원
- 2004년 3월 ~ 현재 : 한국공학대학교(구: 한국산업기술대학교) 전자공학부 교수
- ORCID : <https://orcid.org/0000-0001-5984-0045>
- 주관심분야 : 멀티미디어 신호처리, DTV SOC, 지능형 영상감시 시스템, 딥러닝기반 영상처리