

표준시방서 기반의 의미론적 분석을 반영한 건설 현장 사진 자동 분류 모델 개발

박민건¹ · 김경환^{2*}

¹건국대학교 건축학부 학사과정 · ²건국대학교 건축학부 교수

Development of an Automatic Classification Model for Construction Site Photos with Semantic Analysis based on Korean Construction Specification

Park, Min-Geon¹, Kim, Kyung-Hwan^{2*}

¹Undergraduate Student, Department of Architecture, Konkuk University

²Professor, Department of Architecture, Konkuk University

Abstract : In the era of the fourth industrial revolution, data plays a vital role in enhancing the productivity of industries. To advance digitalization in the construction industry, which suffers from a lack of available data, this study proposes a model that classifies construction site photos by work types. Unlike traditional image classification models that solely rely on visual data, the model in this study includes semantic analysis of construction work types. This is achieved by extracting the significance of relationships between objects and work types from the standard construction specification. These relationships are then used to enhance the classification process by correlating them with objects detected in photos. This model improves the interpretability and reliability of classification results, offering convenience to field operators in photo categorization tasks. Additionally, the model's practical utility has been validated through integration into a classification program. As a result, this study is expected to contribute to the digitalization of the construction industry.

Keywords : Deep Learning, Image Classification, Data Management, Construction Specification

1. 서 론

1.1 연구의 배경 및 목적

최근 4차산업혁명의 진행과 함께 빅데이터 및 인공지능 기술이 폭넓게 도입되면서 많은 산업의 노동생산성이 지속해서 증가하고 있지만, 건설산업의 노동생산성은 매년 감소하는 추세를 보이고 있다(Seong & Yoo, 2022). 또한 ‘산업별 디지털화 수준과 생산성 상관관계’를 살펴보면 건설산업이 타 산업 대비 디지털화 수준이 매우 낮음을 확인할 수 있다(RICON, 2020). 산업의 디지털화 수준을 높여 노동생산성을 증가시키기

위해서는 데이터의 수집을 바탕으로 그 활용도를 높여야 할 것이다. 하지만 건설산업은 ‘낮은 전통산업’의 형태로 경험 의존적이고 참여 주체 간 소통이 원활하지 못하여 데이터의 수집과 활용이 부족한 상황이다(MOLIT, 2018). 이로 인한 활용 가능한 데이터의 부족은 최근 급속한 발전과 함께 여러 산업에서 활용되고 있는 인공지능 기술의 적용도 어렵게 한다(Cheng & Wang, 2018; Kolar et al., 2018). 건설산업의 디지털화를 위한 인공지능을 활용한 연구들에서는 모델의 신뢰도와 성능을 개선하기 위해서 활용 가능한 더 많은 데이터셋 구축의 필요성을 강조하고 있다(Gu et al., 2019; Wang et al., 2022).

한편, 대부분의 건설 사업에서 현장 사진은 흔히 접할 수 있는 건설 현장 데이터이다. 공정, 품질, 안전 등 건설 현장 관리의 중요한 자료가 될 수 있는 현장 사진은 대량으로 보관하고 있으나 이를 체계적으로 분류하고 활용하는 시도는 미미하다. 인공지능 기술을 활용하여 사진을 공종에 따라 자동 분류하는 모델을 개발한 사례(Gil, 2019; Hwang et al., 2019)는 있으나, 의

* Corresponding author: Kim, Kyung-Hwan, Department of Architecture, Konkuk University, Seoul 05029, Korea
E-mail: kykim@konkuk.ac.kr

Received May 22, 2016; revised June 30, 2016
accepted July 1, 2016

미론적인 분석 없이 사용자가 설정한 기준에 따라 분류가 진행되고, 시각적 특성이 유사한 공중에 대해 큰 오차가 발생하는 한계가 있다. 또한 기존의 인공지능 모델 안에서는 어떠한 과정으로 결과가 출력되는지를 알 수 없다는 일명 ‘블랙박스’ 문제가 존재한다(Davide, 2016). 이러한 문제는 인공지능의 해석력을 낮게 만들기 때문에 신뢰도가 떨어진다는 문제로 이어지게 된다.

이에 본 연구에서는 표준시방서의 정보를 반영하여 공중과 건설 현장 내의 객체에 대한 의미론적인 분석이 포함된 새로운 이미지 분류 모델(이하 제안 모델)을 제안하고 이를 활용한 분류 자동화 프로그램을 개발하고자 한다. 이는 건설 현장 사진의 관리와 활용을 용이하게 하고, 건설산업의 디지털화 수준 향상에 기여할 수 있을 것이다.

1.2 연구의 내용 및 절차

본 연구에서는 건설 과정에서 필수적인 문서인 표준시방서로부터 공중별로 건축 자재나 도구의 중요도를 추출하고 이를 사진 내의 객체와 대응시켜 공중별 분류 과정에 반영하는 딥러닝 모델을 개발한다. 이는 기존에 이미지의 특징만을 가지고 분류하는 방식에 비해 해석력이 올라가 성능이 향상되고 부가적으로 객체 탐지가 가능하여 사진을 정리하고 활용하는데 효율성과 생산성이 올라갈 것으로 기대한다. 본 연구의 주요 절차는 다음과 같다.

- 1) 건설 현장 사진과 표준시방서를 공중별로 수집한다.
- 2) 텍스트 마이닝 기법을 통해 표준시방서를 분석하고 사진 내에서 탐지할 건축 자재와 도구를 선정한다. 선정된 객체는 TF-IDF(Term Frequency - Inverse Document Frequency)를 통해 중요도를 산출한다.
- 3) 수집된 건설 현장 사진에서 앞서 선정된 객체들을 라벨링 한다.
- 4) 수집된 데이터를 바탕으로 사용할 딥러닝 모델을 선정하고 모델 구조를 설계한다.
- 5) 선정된 객체를 탐지하는 객체 탐지 모델을 학습하고 학습된 모델을 이미지 분류 모델에 결합한다.
- 6) 완성된 이미지 분류 모델을 학습하여 성능을 측정하고 특징맵 분석과 기존의 분류 모델과의 비교, 분석을 통해 모델의 효용성을 검증한다.

2. 문헌고찰

2.1 컴퓨터 비전

컴퓨터 비전(computer vision)은 사진이나 영상 등의 시각적 정보에서 의미 있는 정보를 추출하여 활용하는

인공 지능 기술의 한 분야이다. 본 연구에서 쓰이는 컴퓨터 비전 분야는 이미지 내의 여러 객체의 위치를 탐지하고, 그 객체를 종류별로 분류하는 객체 탐지 기술(object detection)과 이미지를 사용자가 정한 기준으로 분류하는 이미지 분류 기술(image classification)로 구분할 수 있다(Sandler et al., 2018; Kim et al., 2023b). 이 두 기술은 기본적으로 대표적인 컴퓨터 비전 모델 중 하나인 CNN(Convolutional Neural Network) 모델을 기반으로 하며, Figure 1과 같이 여러 단계를 거쳐 이미지 내의 특징들을 학습하고 추출하는 특징 추출 과정과 이후에 1차원의 데이터로 만들어 분류와 탐지를 진행하는 과정인 분류/탐지 과정을 거쳐 결과를 추출한다(Voulodimos et al., 2018). 여기서 특징은 인간이 물체를 분류할 때 쓰이는 시각적 특징들과 유사하며 예를 들어 직선, 곡선, 색깔과 같이 1차원적인 요소부터 모형, 패턴과 같은 고차원인 요소, 사람이 인식하기 힘든 특징과 같은 추상적인 요소까지도 포함하고 있다.

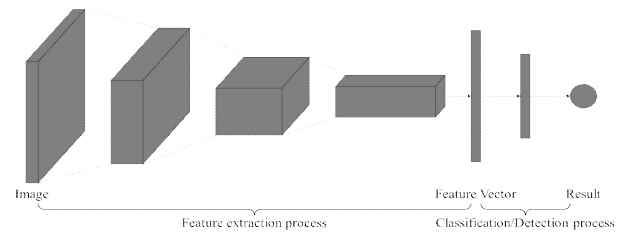


Figure 1. Schematic of image classification model

이와 같은 CNN 모델에서는 특징들을 추출하기 위해 필터를 사용한다. 단계가 깊어질수록 더 다양한 필터를 거쳐 다음 단계로 넘어가 유의미한 특징을 추출하게 되는데, 이때 필터를 거쳐 다음 단계로 넘어가기 전의 결과값을 추출하여 시각화한 것을 특징맵(feature map)이라고 한다(Zhao et al., 2017). Figure 2와 같이 입력층에 가까운 특징맵에서는 선, 모양, 색상과 같이 사람이 이해하기 쉬운 특징들이 나타나게 되고 단계가 깊어질수록 사람이 이해하기 어려운 추상적인 특징들을 추출하게 된다. 따라서 입력층에 가까운 초기 단계의 특징맵을 통해 인공지능 모델이 어떠한 특징을 활용하여 결과를 도출해 내는지 일부 확인할 수 있기에 이미지 내에서 중요한 특징, 모델의 오류를 찾을 수 있으며 이를 통해 모델을 개선할 수 있게 된다(Kim et al., 2023a).

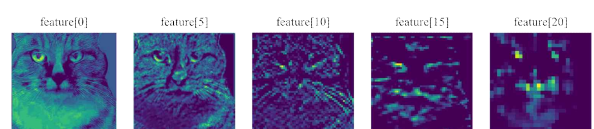


Figure 2. Example of feature map

2.2 건설산업에서의 인공지능

건설 현장의 복잡한 환경으로 인해 건설 현장 사진을 광범위하게 분석하는 것은 어려움이 있다. 따라서 컴퓨터 비전 기술이 사용된 많은 연구들에서는 특정 객체나 작업 등 세부적인 분야에 초점을 맞춘 모델들이 연구되었다. 특히 안전 장비의 탐지 및 착용 여부를 파악하는 안전 관리 연구들이 많이 진행되었다(Nath et al., 2020; Wang et al., 2021; Kim et al., 2023b). 더불어 특정 작업에서의 활용을 위해 Siddula(2016)는 지붕 공사의 위험 요소를 분석하기 위해, Mesfin et al.(2021)은 콘크리트가 타설된 지역을 식별하기 위해 사진을 구역별로 나누는 segmentation 모델을 사용하여 연구를 진행했다.

건설 현장 사진을 공종별로 분류하는 연구로 Gil(2019)에서는 정리되지 않은 사진의 추후 활용을 용이하게 하기 위해 이미지 분류 모델을 적용하였다. 기존에 개발된 여러 딥러닝 모델을 적용하고 그 결과를 비교 분석하여 성능이 높은 모델을 확인하였고, 건설 현장 사진 데이터 세트를 구축하기 위한 기준을 제시하였다. 하지만 해당 연구에서는 연구의 모델이 작업 환경에 대한 의미론적인 분석 없이 연구자가 라벨링한 기준에 따라 학습되어 공종을 분류한다는 문제를 한계점으로 제시한다. 또한 특징맵 분석 결과, 각 공종에서 특징이 되는 요소들이 분류 과정에 혼동을 일으켜 일부 공종의 분류 성능이 저하되었다. 이로 인해 특정 공종 간의 상관관계가 증가하는 문제가 발생하며, 이를 해결하기 위해서 데이터의 수정이 요구된다.

Luo et al.(2018)에서는 건설 현장 사진 내 객체들을 탐지하고 이를 의미론적 및 공간적 관련성을 기반으로 한 관계망을 통해 공종을 분류하는 연구가 진행되었다. 이러한 방식의 모델은 의미론적인 분석이 반영되었기 때문에 신뢰도가 높다. 하지만 사진에서 객체와 객체 간의 거리만을 분류 과정에 반영하였기 때문에 사진에서 탐지되지 못한 객체와 배경은 분류 과정에 제외되게 된다. 따라서 결과가 객체 탐지 성능에 크게 의존한다는 한계점을 가지고 있다.

컴퓨터 비전 기술이 이미지 처리에 중점을 두고 있는 반면, 시방서와 같은 문서 자료를 활용하기 위해서는 텍스트 마이닝 기법이나 자연어 처리 기술을 사용해야 한다. Moon et al.(2022)에서는 텍스트 마이닝 기법과 자연어 처리 기술을 통해 시방서의 검토를 자동화하였고, Akanbi & Zhang(2021)에서는 시방서를 분석하여 건설비용을 추정하는 연구가 진행되었다. 이와 같은 연구들은 전통적인 수동 작업에 비해 시간 및 비용 효율성과 객관성을 갖춘 작업 수행이 가능함을 보여주었다. 이외의 건설산업의 문서정보 활용 인공지능 연구들에

서도 생산성 향상에 중점을 두고 있음을 알 수 있다(Xu, N et al., 2022).

2.3 모델 평가 지표

인공지능 모델들의 성능을 측정하기 위해서 여러 평가지표를 사용하는데, 본 연구에서는 정밀도(precision), 재현율(recall), F1-score 값을 측정하고, 혼동행렬을 통하여 타 공종 간의 혼동 정도를 파악한다. 이 지표들은 실제 값과 예측값에 따른 4가지의 구성 요소로 이루어져 있다. 실제 값과 예측값이 같은 값을 가지는가에 대해 True/False로 나타내며, 예측값이 참값을 가지는가에 따라 Positive/Negative로 나타내어 Table 1과 같이 정리할 수 있다.

Table 1. Confusion matrices and components

| Predicted Class \ Actual Class | Positive | Negative |
|--------------------------------|---------------------|---------------------|
| Positive | True Positive (TP) | False Negative (FN) |
| Negative | False Positive (FP) | True Negative (TN) |

정밀도는 참값으로 예측된 건들 중 실제 참값을 가지는 건의 비율($= \frac{TP}{TP+FP}$), 재현율은 실제 참값을 가지는 건들 중 참값으로 예측된 건의 비율($= \frac{TP}{TP+FN}$)이며, F1-score는 정밀도와 재현율의 조화평균 값($= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$)으로 위 두 개의 평가지표를 적절히 고려한 지표이다. 이러한 지표들을 이용하여 모델의 성능을 측정하고 분석하게 된다(Sokolova & Lapalme, 2009).

2.4 TF-IDF

본 연구에서는 기존의 이미지 분류 모델과 달리 표준 시방서의 정보를 분류 과정에 반영하기 때문에 텍스트 마이닝 기법을 통해 문서의 자연어를 수치화시키는 임베딩 과정을 거치게 된다. 본 연구는 임베딩 방법 중 문서마다의 단어 빈도수를 활용하여 중요도를 부여하는 TF-IDF를 활용하였다. TF-IDF는 식 (1), (2)와 같이 다른 문서에는 많지 않고 특정 문서에 자주 등장하는 정도를 수치화한 지표이다. 이는 특정 문서에서 특정 단어의 출현 빈도를 나타내는 TF(Term Frequency)와 특정 단어가 등장하는 문서의 개수를 나

$$idf(w) = \log\left(\frac{D}{1+df(w)}\right) \quad (1)$$

$$tfidf = tf(w,d) \times idf(w) \quad (2)$$

$tf(w,d)$: 특정 문서 d 에서의 특정 단어 w 의 등장 횟수
 $df(w)$: 특정 단어 w 가 등장한 문서의 수
 D : 문서의 총 개수

타내는 DF(Document Frequency)의 역수 IDF(Inverse Document Frequency)의 곱으로 나타낸다(Park, 2018).

3. 데이터 수집 및 모델 설계, 학습

3.1 데이터 수집

본 연구에서 쓰이는 데이터는 공종별 건설 현장 사진과 표준시방서이다. 공종의 분류는 Gil(2019)의 분류를 따랐으며, 모든 공종을 적용하지 않고 기존 연구에서의 한계점인 작업 간의 혼동을 해결하기 위해 다른 공종과 유사한 작업으로 혼동될 수 있는 6개의 공종들을 선택하여 연구를 진행했다.

먼저 건설 현장 사진은 Gil(2019)에서 제안하는 이미지 수집 가이드라인을 따르며 모두 검색엔진의 이미지 검색을 통해 수집하였다. 공종별로 약 50장씩 총 307장의 사진을 수집하였고, 사진의 데이터 세트 구성은 Table 2와 같다. 수집한 건설 현장 사진으로는 객체 탐지 모델을 학습하기에 객체의 데이터가 부족하여 단일 객체의 사진 50장을 추가로 수집하였다.

표준시방서는 국가법령정보센터에서 내려받아 사용하였으며 공종별로 세부 공사 기준서를 모두 병합하여 총 6개의 공종 파일로 나누어 사용하였다.

Table 2. Configuration of dataset for image classification

| Category | Total | Train | Test |
|-------------|-------|-------|------|
| carpentry | 50 | 32 | 18 |
| earthwork | 55 | 38 | 17 |
| landscaping | 51 | 35 | 16 |
| masonry | 50 | 35 | 15 |
| painting | 51 | 37 | 14 |
| plastering | 50 | 37 | 13 |
| Total | 307 | 214 | 93 |

3.2 표준시방서 정보 추출

기존의 문서 정보 활용 연구에서는 작업 생산성 향상을 목적으로 두지만, 본 연구에서는 해석 가능성이 떨어지는 딥러닝 모델의 신뢰도를 높이기 위한 목적으로 표준시방서가 사용되었다. 표준시방서를 통하여 공종별로 중요한 객체를 선정하며, 해당 객체가 얼마나 자주 언급되는지를 분석하여 각 공종의 중요도를 산출한다. 이러한 방식으로 추출된 객체의 중요도는 이후 공종을 분류하는 과정에 반영되어 중요도가 높은 공종에 가중치를 주게 된다.

3.2.1 탐지 객체 선정

사진 내의 객체를 공종별 분류 과정에 반영하기 위해 탐지할 객체를 먼저 선정해 주어야 한다. 이를 위해 파

이썬 라이브러리 중 하나인 'KoNLPY'라는 한글 형태소 분석기를 활용하여 각 공종의 세부 공사 기준서에서 명사만을 추출하였다. 이후 빈도수로 내림차순으로 정렬하여 사진 내에서 탐지할 수 있는 건축자재와 도구에 해당하는 단어를 선정하였고, 그중에서 비슷한 객체를 의미하는 단어끼리 묶었다. 또한 표준시방서에서 선정되지 않은 단어일지라도 사진 내에서 공종을 특징짓는 데 영향을 줄 수 있는 건축자재나 도구 또한 탐지할 객체로 선정하였다. 이러한 객체는 표준시방서에서 선정된 객체 중에서 중요도가 비슷하다고 생각되는 객체의 빈도수를 갖는다고 가정하고 설정하였다. 결과적으로 탐지할 객체는 총 20개로 선정하여 빈도수를 추출하였다.

3.2.2 TF-IDF 추출

TF-IDF를 추출하기 위해 선정된 객체의 빈도수를 사용해야 한다. 하지만 토공사 관련 표준시방서는 총 48개의 세부 공사 기준서가 존재하는 반면, 목공사는 세부 공사 기준서가 총 5개 밖에 존재하지 않는다. 이러한 데이터양의 차이는 결과가 데이터가 많은 쪽으로 쏠리는 편향의 문제가 발생할 수 있다. 따라서 각 공종에 대하여 빈도수의 합을 나눠주는 정규화를 진행한 뒤에 TF-IDF를 추출하였다. 그 결과, Table 3과 같이 총 14개의 category와 6개의 공종에 대한 TF-IDF 결과를 추출하였다.

3.3 딥러닝 모델 구조 설계

본 연구의 제안 모델은 사진 내 객체에 대한 의미론적인 분석을 분류 과정에 반영하기 위하여 Figure 3과 같이 모델 구조를 설계하였다. 객체 탐지 모델과 표준시방서로부터 추출한 객체에 대한 공종별 중요도와 특징 추출 과정에서 나온 이미지 내의 특징들을 결합하여 분류 과정을 거치는 구조이다. 객체 탐지 모델로 대외적으로 성능이 입증된 'YOLO'모델을 사용하고, 이미지 분류 모델에서의 이미지 내 특징 추출을 위해서는 Gil(2019)에서 성능이 가장 높게 나온 'Mobilenet' 모델의 구조를 사용하였다(Redmon et al., 2016). 이후 공종을 분류하는 과정에서는 기본적인 DNN 구조인 Fully Connected Layer를 사용하였다. 전체적인 모델의 진행 과정은 다음과 같이 진행된다.

(1) 건설 현장 이미지를 Mobilenetv2의 구조를 통해 이미지 내의 특징을 추출한다.

(2) 객체 탐지 모델에서는 선정된 총 20개의 객체를 탐지한 후, 총 14개의 카테고리 묶어 객체의 개수를 추출한다.

(3) 추출된 객체의 개수는 객체별로 만들어진 TF-IDF 테이블에 곱해진다. 즉, 특정 객체가 탐지되지

Table 3. TF-IDF table of object-category

| Object | earthwork | painting | carpentry | plastering | landscaping | masonry |
|---|-----------|----------|-----------|------------|-------------|---------|
| board | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 |
| brick; brick & bricks | 0.22 | 0.00 | 0.00 | 0.00 | 0.20 | 0.86 |
| concrete; concretebag & basin | 0.45 | 0.07 | 0.00 | 0.92 | 0.42 | 0.49 |
| excavator | 0.67 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 |
| liquid_material; plasticbarrel & square can | 0.00 | 0.99 | 0.00 | 0.10 | 0.00 | 0.00 |
| long hand float | 0.00 | 0.00 | 0.00 | 0.17 | 0.00 | 0.00 |
| pile | 0.28 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| plant | 0.00 | 0.00 | 0.00 | 0.00 | 0.57 | 0.00 |
| plastering_tool; trowel & rainboot & plasterer shoe | 0.00 | 0.00 | 0.00 | 0.31 | 0.00 | 0.00 |
| roller | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| steel | 0.20 | 0.08 | 0.00 | 0.00 | 0.00 | 0.14 |
| stud | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.00 |
| truck; dumptruck & mixer truck | 0.67 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| wood | 0.00 | 0.05 | 0.97 | 0.00 | 0.25 | 0.00 |

않았다면 해당 객체의 TF-IDF는 0이 되고, 탐지되었다면 해당 객체의 TF-IDF는 객체가 탐지된 개수의 배수만큼 커지게 된다.

(4) 개수가 곱해진 TF-IDF 테이블은 일차원의 데이터로 변환된 뒤 이미지 특징 추출 과정에서 나온 데이터와 결합한다.

(5) 결합한 데이터로 분류 과정을 거쳐 결과를 나타낸다.

3.4 딥러닝 모델 학습

본 연구에서는 기존의 이미지 분류 모델, 제안 모델에 활용하는 객체 탐지 모델, 그리고 최종적으로 제안하는 이미지 분류 모델, 총 3개의 모델을 학습한다.

3.4.1 객체 탐지 모델

객체 탐지 모델은 'YOLO v8 nano'를 사용하였고 수집한 건설 현장 사진과 객체 단일 사진, 총 357장의 사진에서 총 1,646개의 객체를 학습시켰다. 객체 탐지 모델은 모든 데이터를 학습데이터에 포함하여 학습시

켜 객체 탐지 모델의 성능을 강제로 높여 사용했는데, 이러한 방식은 객체에 대한 데이터 부족으로 인해 선택하였다. 결과적으로 제안 모델의 객체 탐지 모델 성능은 평균 정밀도(mAP50) 0.89에 도달하였다. 본 연구의 주목적은 제안 모델이 표준시방서의 정보를 잘 반영하여 분류 성능에 향상과 효용성이 있는가를 입증하기 위한 연구이며, 추후에 객체에 대한 충분한 데이터가 쌓여 이를 보완한다면 객체 탐지 모델의 성능은 충분히 확보할 수 있을 것이다.

이렇게 학습된 모델을 활용하여 사진 내의 객체를 탐지하고 이 객체들을 앞서 설정한 14개의 카테고리 분류하여 개수를 추출한다. 추출된 데이터는 TF-IDF 테이블과 곱해져 새로운 데이터로 만들어지고 1차원의 데이터로 변환 후 이미지 분류 모델에 사용된다.

3.4.2 이미지 분류 모델

이미지 분류 모델은 연구의 대조군인 기존의 이미지 분류 모델과 본 연구의 제안 모델, 2가지를 학습한다. 두 모델 모두 총 307장의 사진을 학습데이터와 테스트

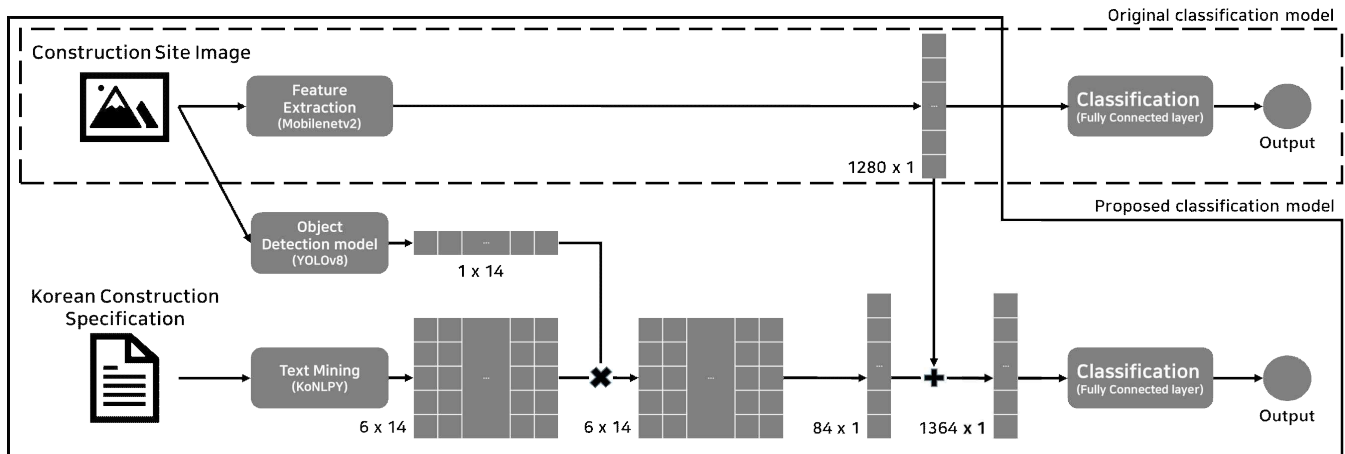


Figure 3. Schematic of the proposal model

데이터로 나누어 0.7:0.3의 비율로 학습하였다. 다만, 공종별 50개의 데이터는 학습하기에 상대적으로 적은 양의 데이터이다. 따라서 적은 데이터 양의 문제를 해결하기 위해 ImageDataGenerator 라이브러리를 사용하여 학습 데이터를 증식¹⁾ 하였고, 학습 데이터에 과적합하는 현상을 줄이기 위해 dropout 기법²⁾을 적용하였다.

기존의 이미지 분류 모델은 제안 모델과 비교하기 위한 기존에 일반적으로 사용되는 구조의 이미지 분류 모델이다. Figure 3의 original classification model 구조와 같이 이미지만을 가지고 Mobilnetv2의 구조로 이미지의 특징을 추출하고 이를 공종별로 분류하는 구조를 가진다.

본 연구의 제안 모델은 기존의 이미지 분류 모델과 같이 이미지 내의 특징을 Mobilnetv2의 구조를 통해 추출하고, 이를 객체 탐지 모델과 표준시방서를 활용하여 추출된 정보와 함께 공종을 분류하는 과정을 거치게 된다. 최종 분류 과정에서는 총 6개의 공종마다의 확률값으로 나타내게 되고, 이후 가장 큰 확률값을 가지는 공종을 최종 결과값으로 선택하게 된다.

4. 학습 결과 및 분석

4.1 학습 결과

먼저 기존 이미지 분류 모델과 제안 모델의 테스트 데이터에 대한 결과는 Table 4와 같이 각각 75%, 81%의 정확도를 보였으며 혼동행렬의 값은 Figure 4와 같다. 종합적으로 보았을 때 본 연구에서 제안하는 모델이 더 많은 공종에서 높은 성능을 보여주고 있다.

Table 4. Comparison of model results

| Category | Previous model f1-score | Proposal model f1-score | Images |
|---------------|-------------------------|-------------------------|--------|
| carpentry | 0.77 | 0.91 | 18 |
| earthwork | 0.89 | 0.83 | 17 |
| landscaping | 0.87 | 0.80 | 16 |
| masonry | 0.78 | 0.86 | 15 |
| painting | 0.50 | 0.64 | 14 |
| plastering | 0.61 | 0.75 | 13 |
| avg. accuracy | 0.75 | 0.81 | 93 |

분류 결과값과 혼동행렬을 살펴보았을 때 미약하지만 전반적인 공종끼리의 혼동되는 정도가 줄어든 것을 확인할 수 있었다. 그중 미장공사와 도장공사의 혼동되는 결과들이 많이 보였는데 이는 Gil(2019)에서 언급되었

던 것과 같이 미장공사와 도장공사 두 공종 간의 상관관계가 크다는 것을 확인할 수 있었다.

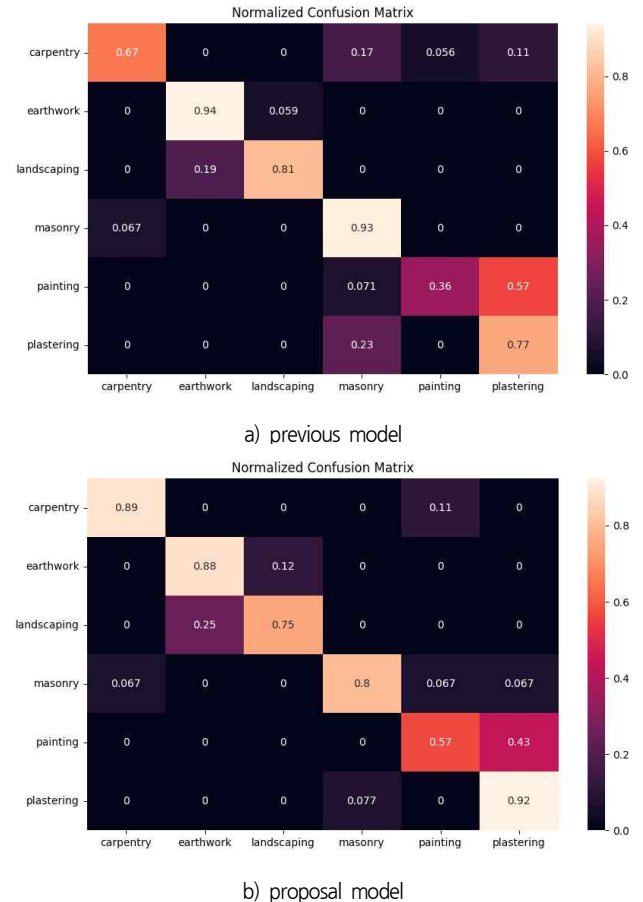


Figure 4. Confusion matrix of model results

4.2 결과 분석

4.2.1 특징맵 분석

기존 모델과 제안 모델의 특징맵을 각각 추출한 결과 두 모델의 특징맵은 동일하게 추출되었다. 이는 이미지의 특징을 추출하는 과정과 결과는 동일한 것이므로 두 모델의 성능 차이는 분류 과정에서 결정되는 것을 의미한다. 따라서 분류 과정에 추가로 들어간 표준시방서에서 추출한 공종에 따른 객체별 가중치가 모델의 성능을 향상시켰다고 해석할 수 있다.

특징 추출 과정에서 어떠한 시각적인 특징을 추출하여 결과를 도출하는지 확인하기 위해 특징맵을 시각화하여 살펴보았다. Figure 5는 도장공사가 진행 중인 사진과 특징맵을 시각화한 것이다. 첫 번째 특징맵은 선, 색상 등과 같이 낮은 수준의 특징이 추출되기에 전반적인 객체의 모습이 뚜렷하게 추출되었으며 단계가 깊어질수록 고수준의 특징이 추출되어 세 번째 특징맵에서는 작업자가 들고 있는 롤러 부분만을 특징으로 추출하였다고 해석할 수 있다. 이후의 특징맵들은 사람이 이해하기는 어려운 추상적인 특징만을 추출하였기에 분석하기엔 어려움이 있었다.

1) 이미지를 회전시키거나, 크기를 조절하는 등 기존의 학습 데이터를 변형하여 가상적으로 데이터의 양을 늘려 모델의 과적합을 줄이는 데 도움을 주는 기법이다(Shorten, 2019).

2) 학습 과정 중에 신경망의 무작위 뉴런을 비활성화하여 모델이 학습 데이터에 과도하게 의존하는 것을 방지하고, 결과적으로 일반화 능력이 향상된 강건한 학습을 가능하게 해준다(Srivastava, 2014).

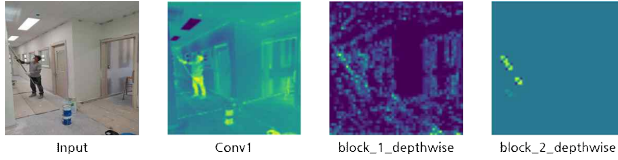


Figure 5. Feature maps for paintwork photo

Figure 6은 여러 공종의 사진들이며 동일한 필터에서 나온 각 사진의 특징맵들이다. 해당 필터는 사람이나 기계와 같은 형태가 뚜렷한 객체보다는 이미지의 전반적인 배경을 특징으로 추출하고 있다. 다만 건설 현장의 복잡하고 다양한 배경은 특징을 추출하는 데 있어 노이즈로 작용할 수 있으며 기존의 결과와 같이 타 공종과 혼동되는 요소들로 인해 공종끼리의 상관관계가 높아져 분류 성능에 영향을 줄 수 있다. 하지만 토공사와 조적공사, 목조공사와 미장공사의 작업 배경이 다르듯이 배경 또한 공종을 분류하는 데 있어서 중요한 요소로 작용한다. 이러한 이유로 Luo et al.(2018)의 연구와 같이 건설 현장의 배경을 제외하고 단순히 객체만으로 공종을 분류하는 것은 객체에 대한 의존도가 높아져 분류 성능이 낮아지는 결과가 있을 수 있다.

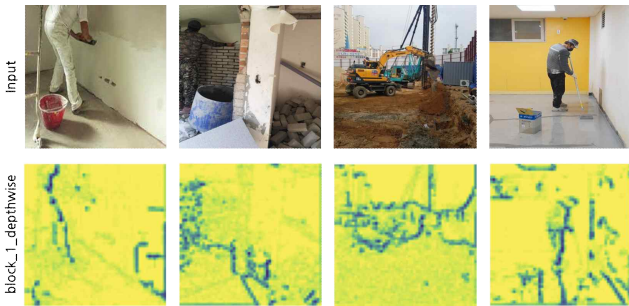


Figure 6. Feature map by category

다만 이와 같은 특징맵 분석은 주관적인 해석이 들어가기 때문에 분석의 신뢰도가 떨어지는 문제가 있다. 그럼에도 불구하고 분류 초기의 특징 추출 과정을 살펴보면 모델의 특징 추출 과정과 분류 과정의 방향성을 예상해 볼 수 있다는 점에서 특징맵 분석은 의미가 있다. 이러한 분석을 통해 Figure 4와 같이 혼동행렬의 값들이 향상된 이유는 분류 과정에 반영된 공종에 따른 객체별 가중치 때문임을 알 수 있고, 이는 기존 이미지 분류 모델이 가지고 있던 혼동되는 요소로 인해 공종끼리 상관관계를 가진다는 한계점을 본 연구의 제안 모델이 일부 해결하였다고 볼 수 있다. 또한 건설 현장의 배경이 객체에 대한 높은 의존도를 해결하는 데 있어서 중요한 요소임을 확인할 수 있다.

4.2.2 케이스 분석

표준시방서에서의 객체별 가중치가 모델에 반영된 것을 확인하기 위해 미장공사 사진을 올바르게 예측한

테스트 데이터를 분석하였다. Figure 7은 실제로는 미장공사의 사진이지만 기존 이미지 분류 모델은 조적공사로 예측하였다. 제안 모델에서 객체 탐지 결과로 흙손 1개, 장화 3개, 플라스틱 통 1개가 탐지되었고 공종별 확률은 Table 5와 같다. 기존 이미지 분류 모델의 미장공사의 확률은 0.3759, 본 연구에서 고안한 모델의 미장공사의 확률은 0.9782로 확률값이 약 2배 이상 높아진 것을 확인할 수 있었다. 반면 기존 이미지 분류 모델이 조적공사로 잘못 예측한 것은 흙손을 사용하여 시멘트를 바르는 조적공사와 미장공사의 비슷한 작업 모습으로 인한 오류라고 해석해 볼 수 있다.



Figure 7. Object detection result for plastering photo

Table 5. Comparison probability value and result for Figure 7

| Category | Original Model | Proposal Model |
|-------------|----------------|----------------|
| carpentry | 0.0019 | 0.0011 |
| earthwork | 0.0025 | 0.0003 |
| landscaping | 0.0097 | 0.0014 |
| masonry | 0.5980 | 0.0160 |
| painting | 0.0120 | 0.0030 |
| plastering | 0.3759 | 0.9782 |
| Result | masonry | plastering |

하지만 모든 데이터가 위와 같이 분류 성능이 올라간 것은 아니며 일부는 성능이 저하되기도 하였다. 분류 성능이 저하된 공종을 살펴본 결과, 여러 공종에 중요도를 갖는 객체가 오히려 혼동 요소로 작용함을 예상해 볼 수 있었다. excavator가 탐지된 landscaping 사진의 경우, earthwork로 분류되는 경우가 많은 것으로 보아 excavator가 earthwork와 landscaping의 혼동 요소로 작용하고 결과적으로 landscaping에 대한 분류 성능이 떨어졌다고 할 수 있다. 이는 TF-TDF로 공종에 가중치를 주었지만, 실제 학습을 할 때 사용되는 모델 내의 가중치는 직접 제어할 수 없기 때문이라고 해석해 볼 수 있다.

그럼에도 Table 6과 같이 탐지된 객체가 4개 이상인 데이터의 정확도는 4개 미만인 데이터의 정확도보다

8%가량 높게 측정되었다. 따라서 더 많은 객체를 탐지하여 분류 과정에 더 많은 정보를 반영할 수 있다면 객체가 공중을 분류하는데 혼동을 준다는 문제를 해결할 수 있고 더불어 정확도의 향상도 기대할 수 있다.

Table 6. Accuracy based on the number of objects detected

| Number of objects detected | Accuracy | Images |
|----------------------------|----------|--------|
| 4≤ | 0.85 | 46 |
| <4 | 0.77 | 47 |

4.3 분류 자동화

제안 모델에서 나온 여러 정보를 저장하고 사진을 공중별 파일로 이동시켜 분류하는 프로그램을 제작하였다. Table 7과 같이 공중 분류 결과, 사진의 시간 정보, 분류된 파일 위치, 객체의 정보를 저장하게 되어 이후 사용자가 원하는 객체와 시간, 공중을 검색하며 사진을 찾을 수 있다.

5. 결론

IT와 산업 간의 연계, 인공지능의 발전 등으로 노동생산성이 증가하고 있는 타 산업과 달리 건설산업은 낮은 디지털화 수준과 함께 노동생산성이 지속적으로 낮아지고 있다. 본 연구에서는 건설산업의 디지털화 수준을 끌어올리고자 디지털화의 기반이 되는 데이터 수집에 초점을 맞추어 수집된 건설 현장의 사진이 공중별로 분류되어 이후에 다시 활용될 수 있도록 새로운 이미지 분류 인공지능 모델을 개발하였다. 또한 기존 공중 분류 인공지능 모델의 한계점들을 극복하고자 실제 공사에 사용되는 표준시방서의 정보를 분류 과정에 반영하여 개발하였으며 학습 결과를 비교, 분석하였다. 최종적으로 사진 분류를 자동화하는 프로그램을 만들어 실제 활용 시의 효용성을 확인하였다.

본 연구에서 제안하는 모델은 기존 이미지 분류 모델보다 약 5~6%의 성능 향상이 이루어졌고, 특징맵 분석을 통해 분류 과정에 반영한 표준시방서의 정보가 성능 향상에 기여했음과 건설 현장의 배경이 공중을 분류하는 데 있어 중요한 특징임을 확인하였다.

다만 데이터의 부족으로 인하여 객체 탐지 모델에서 학습 데이터를 그대로 이용하여 성능을 인위적으로 향상시켰다는 점, 분류 과정에 반영할 객체를 선정할 때 일부 주관적인 작업이 들어갔다는 점에서 한계점을 가지고 있다. 그럼에도 불구하고 본 연구에서의 중요한 기여점은 다음과 같다.

1) 기존에 존재했던 컴퓨터 비전 활용 연구들과 달리 본 연구와 같이 표준시방서와 같은 문서정보를 이미지를 분류하는 과정에 반영하는 연구는 찾아볼 수 없었다. 하지만 최근에 주목을 받고 있는 여러 형태의 데이터를 동시에 처리하는 멀티 모달 인공지능의 일부라고 할 수 있기에 본 연구와 같은 모델의 개발은 앞으로 더욱 많은 연구에 활용될 것으로 기대한다. 또한 이러한 방식은 해석가능한 문서 정보가 분류 과정에 반영되므로 기존의 의미론적인 분석 없이 진행되던 공중 분류에 해석력이 부여된다고 할 수 있다.

2) 객체 탐지 모델이 결합되어 있는 분류 모델이므로 단순히 이미지를 공중만으로 분류하는 것이 아니라 이미지 내의 객체를 탐지하기 때문에 이들 데이터를 활용하여 실무자의 사진 관리에 용이성과 효율성을 높여 줄 수 있을 것이다.

본 연구는 단일 공중에 대해서만 분류를 진행했으나, 이러한 접근은 다중 공중을 분류하기 위한 기초를 마련한다. 향후 연구에서 여러 태그나 레이블을 할당할 수 있는 auto-tagging 알고리즘을 활용하거나, 방대한 데이터를 학습한 GPT 모델을 사용한다면 더 넓은 공중들을 더욱 정확하게 분류할 수 있을 것이다.

이러한 사진 분류 인공지능 연구로 높은 분류 성능과 다양한 기능을 갖춘 시스템이 개발된다면 현장 기사들의 사진 수집과 관리가 용이하게 할 것이다. 또한 이를 데이터 관리 플랫폼, 클라우드를 연결하여 건설산업의 다양한 이해관계자들이 제약 없이 주고받도록 한다면 건설산업의 디지털화 수준과 생산성은 더욱 향상될 것이다. 이러한 관점에서 본 연구는 앞으로 경험 중심적인 건설산업에서 데이터 중심적인 건설산업으로 변화되면서 중요해지는 데이터 수집과 활용에 기여할 것으로 기대한다.

Table 7. Structure of data storage

| | result_image_path | Class | date_made | date_taken | moved_image_path | basin | board | .. | wood | Trowel |
|----|--|------------|---------------------|---------------------|---|-------|-------|----|------|--------|
| 3 | C:\Users\W***\Desktop\Wtest\Wtest_image\Wcarpentry (17).jpeg | carpentry | 2023:10:11 13:56:04 | 2018:05:03 16:57:36 | c:\Users\W***\Desktop\Wtest\Wclassified_images\Wcarpentry\Wcarpentry (17).jpeg | 0 | 0 | .. | 0 | 0 |
| 17 | C:\Users\W***\Desktop\Wtest\Wtest_image\Wcarpentry (6).jpeg | painting | 2023:10:11 13:56:04 | | c:\Users\W***\Desktop\Wtest\Wclassified_images\Wpainting\Wcarpentry (6).jpeg | 0 | 5 | .. | 2 | 0 |
| 59 | C:\Users\W***\Desktop\Wtest\Wtest_image\Wmasonry (32).jpeg | masonry | 2023:10:11 13:56:04 | 2018:04:16 09:57:04 | c:\Users\W***\Desktop\Wtest\Wclassified_images\Wmasonry\Wmasonry (32).jpeg | 0 | 0 | .. | 0 | 0 |
| 83 | C:\Users\W***\Desktop\Wtest\Wtest_image\Wplastering (2).jpeg | plastering | 2023:10:11 13:56:04 | | c:\Users\W***\Desktop\Wtest\Wclassified_images\Wplastering\Wplastering (2).jpeg | 1 | 0 | .. | 0 | 0 |

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1F1A1063286).

References

1. Akanbi, T., & Zhang, J. (2021). "Design information extraction from construction specifications to support cost estimation." *Automation in Construction*, 131, 103835.
2. Castelveccchi, D. (2016). "Can we open the black box of AI?" *Nature News*, 538(7623), pp. 20–23.
3. Cheng, J. C., & Wang, M. (2018). "Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques." *Automation in Construction*, 95, pp. 155–171.
4. Gil, D. Y. (2019). "Automated method to classify construction site photos based on deep learning." Master Dissertation, Yonsei University.
5. Gu, H. M., Seo, J. H., & Choo, S. Y. (2019). "A Development of Facade Dataset Construction Technology Using Deep Learning-based Automatic Image Labeling." *Journal of the architectural institute of Korea planning & design*, 35(12), pp. 43–53.
6. Hwang, J., Kim, J., Chi, S., & Seo, J. (2019). "Automated Training Database Development through Image Web Crawling for Construction Site Monitoring." *Journal of Civil and Environmental Engineering Research*, KSCE, 39(6), pp. 887–892.
7. Kim, H., Ahn, J., Lee, T. Y., & Choi, B. (2023a). "The Object Detector for Aerial Image Using High Resolution Feature Extractor and Attention Module." *The Journal of Korean Institute of Communications and Information Sciences*, 48(1), pp. 1–11.
8. Kim, K., Kim, K., & Jeong, S. (2023b). "Application of YOLO v5 and v8 for Recognition of Safety Risk Factors at Construction Sites." *Sustainability*, 15(20), 15179.
9. Kolar, Z., Chen, H., & Luo, X. (2018). "Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images." *Automation in Construction*, 89, pp. 58–70.
10. Luo, X., Li, H., Cao, D., Dai, F., Seo, J., & Lee, S. (2018). "Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks." *J. Comput. Civ. Eng.*, 32(3), 04018012.
11. Mesfin, W. M., Cho, S., Lee, J., Kim, H. K., & Kim, T. (2021). "Deep-learning-based segmentation of fresh or young concrete sections from images of construction sites." *Materials*, 14(21), 6311.
12. MOLIT. (2018). "Smart Construction Technology Roadmap." Ministry of Land, Infrastructure(MOLIT).
13. Moon, S., Lee, G., & Chi, S. (2022). "Automated system for construction specification review using natural language processing." *Advanced Engineering Informatics*, 51, 101495.
14. Nath, N. D., Behzadan, A. H., & Paal, S. G. (2020). "Deep learning for site safety: Real-time detection of personal protective equipment." *Automation in Construction*, 112, 103085.
15. Park, D. S., & Kim, H. J. (2018). "A Proposal of Join Vector for Semantic Factor Reflection in TF-IDF Based Keyword Extraction." *The Journal of Korean Institute of Information Technology*, 16(2), pp. 1–16.
16. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
17. RICON FOCUS. (2020). "Innovation plan for the construction industry according to the acceleration of the 'digital economy'." Korea Research Institute For Construction Policy(RICON).
18. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). "Mobilenetv2: Inverted residuals and linear bottlenecks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520.
19. Seong, Y., & Yoo, W. (2022). "Korean construction industry productivity analysis." Construction & Economy Research Institute of Korea.
20. Shorten, C., & Khoshgoftaar, T. M. (2019). "A survey on image data augmentation for deep learning." *Journal of big data*, 6(1), pp. 1–48.
21. Siddula, M., Dai, F., Ye, Y., & Fan, J. (2016). "Unsupervised feature learning for objects of interest detection in cluttered construction roof site images." *Procedia Engineering*, 145, pp. 428–435.
22. Sokolova, M., & Lapalme, G. (2009). "A systematic analysis of performance measures for classification tasks." *Information processing & management*, 45(4), pp. 427–437.
23. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research*, 15(1), pp. 1929–1958.
24. Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). "Deep learning for computer vision: A brief review." *Computational intelligence and neuroscience*, 2018.
25. Wang, Z., Kim, J. S., Ham, N. H., & Kim, J. J. (2022). "Proposal and Verification of the Faster

- R-CNN Regarding the Worker and Machine Interference Scope Detection Model to Prevent On-site Safety Accidents.” *Journal of the Architectural Institute of Korea*, 38(4), pp. 217–228.
26. Wang, Z., Wu, Y., Yang, L., Thirunavukarasu, A., Evison, C., & Zhao, Y. (2021). “Fast personal protective equipment detection for real construction sites using deep learning approaches.” *Sensors*, 21(10), 3478.
27. Xu, N., Zhou, X., Guo, C., Xiao, B., Wei, F., & Hu, Y. (2022). “Text Mining Applications in the Construction Industry: Current Status, Research Gaps, and Prospects.” *Sustainability*, 14(24), 16846.
28. Zhao, Z., Fan, X., Xu, G., Zhang, L., Qi, Y., & Zhang, K. (2017). “Aggregating deep convolutional feature maps for insulator detection in infrared images.” *IEEE Access*, 5, pp. 21831–21839.

요약 : 4차 산업 시대에서의 데이터는 산업의 생산성을 높이는 데 매우 중요한 역할을 하고 있다. 활용 가능한 데이터가 부족한 건설산업의 디지털화 수준을 높이기 위해서 본 연구에서는 건설 현장 사진을 공중별로 분류하는 모델을 연구하였다. 이 미지만을 가지고 분류하는 기존의 이미지 분류 모델과 달리, 본 연구는 표준시방서에서 객체와 공중 간의 중요도를 추출하여 이를 분류 과정에 반영하는 방식으로 공중에 대한 의미론적인 분석을 포함한 분류 모델을 제안하였다. 객체와 공중 간의 중요도는 사진 내에서 탐지한 객체와 표준시방서의 정보를 연결하여 추출한 후 모델에 반영하였고, 이러한 방식으로 개발된 모델을 분류 프로그램에 적용하여 실제 실무에서의 유용성을 확인해 보았다. 제안한 모델은 결과에 해석가능성과 신뢰도를 높여주는 것뿐만 아니라 현장 기사들이 사진을 분류하는데 용이성을 주게 되며, 이러한 연구의 결과는 건설산업의 디지털화에 기여할 수 있을 것이다.

키워드 : 딥 러닝, 사진 분류, 데이터 관리, 표준 시방서
