

건설현장 컴퓨터비전 AI 성능에 대한 학습 이미지 데이터셋 크기 및 특화성의 영향 분석*

Investigating the Effects of Training Image Dataset's Size and Specificity on Visual Scene Understanding AI in Construction

김진우** · 지석호***

Jinwoo Kim** · Seokho Chi***

Abstract

Visual scene understanding AI, a pivotal factor for digital transformation and robotic automation in construction, has primarily been researched under the hypothesis that the more training images, the higher the model performance. Alternatively, one can hypothesize that prioritizing activity-specific training images tailored to each construction phase would be more critical than merely enlarging the size of the dataset. This approach is particularly vital in dynamic construction environments where visual characteristics undergo significant changes across the construction phases, from earthmoving, foundation, and superstructure to finishing activities. In this background, we investigate the effects of a training image dataset's size and specificity on visual scene understanding AI in construction. We build an all-in-one, universal training image dataset as well as an activity-specific dataset, varying the number of training images. We then train vision-based worker detection models using each dataset and assess their performance in activity-specific, dynamic test environments. We analyze the optimal performance achieved in each test environment and how the model's performance varies depending on the dataset's size over the entire test phase. Our findings will help scientifically validate the dual hypotheses and lay a solid foundation for building and updating a training image dataset when developing a visual scene understanding AI model in dynamic construction sites.

Keywords: Construction Site, Visual Scene Understanding, AI, Training Dataset, Activity-Specific

1. 서론

건설산업은 디지털 전환과 로봇 자동화를 통해 생산성 및 안전성 혁신을 도모하고 있다. 미래 현장에서는 (반)자율로봇이 육체적으로 힘든 작업을 주로 수행하는 반면, 사람은 이 로봇을 운영하거나 새로

운 작업방식을 제안하는 등 고부가가치의 업무를 맡게 될 것이다. 더불어, 무인비행체 등 이동식 로봇은 대규모 작업장을 자율적으로 돌아다니면서 현장정보를 체계적으로 수집-저장-디지털화할 수 있고, 이를 통해 건설현장을 효율적으로 모니터링하고 관리할 수 있다. 이처럼 최첨단 디지털 및 로봇 기술은 건

*이 논문은 한양대학교 교내연구지원사업으로 연구되었음(HY-20240000001265).

**한양대학교 건설환경공학과 조교수(주저자·교신저자: jinwookim@hanyang.ac.kr)

***서울대학교 건설환경공학부 교수(shchi@snu.ac.kr)

설산업의 고질적인 문제였던 생산성 및 안전성을 혁신하는 데 큰 도움이 될 것이다.

디지털 전환과 로봇 자동화를 실현하기 위해서는 현장에서 수집되는 다량의 이미지를 분석하여 끊임 없이 변화하는 현장상황을 이해할 수 있는 컴퓨터비전 AI가 필수적이다. 예를 들어, 건설현장에서 (반)자율로봇이 돌아다니기 위해서는 주변의 작업자, 구조물, 이동통로 등을 인식하고 적절한 의사결정을 수행해야 한다. 이에 따라 수많은 연구자들이 물체 탐지 및 추적, 단위분할, 행동인식, 자세추정 등 다양한 컴퓨터비전 AI 기술을 조사하고, 이를 건설현장에 활용할 수 있는 형태로 개선-맞춤개발해왔다.

상당한 연구개발이 이루어졌지만 기존의 컴퓨터비전 AI 기술은 실무 적용성이 매우 낮아 실제 현장에서는 제대로 활용되지 못하고 있다. 가장 큰 원인 중 하나는 건설현장의 상황이 토공-기초-골조-마감 작업에 걸쳐 시시각각 변한다는 점을 고려하지 않고, 단순히 “학습 이미지 데이터 수가 많을수록 AI 성능이 증가할 것이다”라는 가설을 토대로 전 공종을 포괄하는 하나의 범용 데이터셋을 구축했기 때문이다. AI 문제로 변환해서 설명하면, 실제 건설현장인 시험 데이터셋의 분포가 동적으로 계속 변함에도 불구하고, 기존 연구에서는 초기에 한번 구축한 이후로 전혀 업데이트하지 않는, 정적이고 고정된 방식을 적용했다는 것이다.

이런 맥락에서 “각 공종에 특화된 학습 이미지 데이터셋을 지속적으로 업데이트-구축하면 더 높은 성능의 AI 모델을 학습할 수 있다”라는 기존과 대립하는 가설을 세울 수 있다. 이 두 가지 모두 합리적인 가설이지만, 건설분야에서 학습 이미지 데이터셋의 크기와 특화성(각 공종에 특화된 정도)이 컴퓨터비전 AI 성능에 미치는 영향에 대한 정량적 근거나 자료가 부족하여 어떤 방식으로 학습 이미지셋을 구축해야 하는지 미지수로 남아있다. 이에 따라 본 연구는 학습 이미지 데이터셋의 크기를 변경해가며 공종별

특화 데이터셋과 전 공종 범용 데이터셋을 구축하고 컴퓨터비전 AI 모델을 학습한 뒤, 토공-기초-골조-마감공종 전 단계에서 그 성능을 차례대로 비교분석한다. 특히, 건설사업에서 가장 중요한 자원이고 컴퓨터비전 AI의 기반이 되는 작업자 인식을 연구범위로 한정하였다.

2. 선행연구

2.1 건설현장 컴퓨터비전 AI 연구

최근 몇 년간 심층학습 분야의 큰 발전과 함께 많은 연구자들이 건설현장 이미지를 분석하는 방법들을 고안해왔다. 일부는 작업자의 작업유형을 인식하기 위해 합성곱신경망, 순환신경망 등 다양한 유형의 심층학습 구조를 활용하였으며(Luo et al., 2020; Roberts et al., 2020), 다른 연구자들은 굴삭기, 덤프트럭 등 중장비에 대한 그 가능성을 검증하였다(Kim et al., 2018; Kim and Chi, 2019). 최근에는 토공-기초-골조-마감공사 건설사업의 전 생애주기에 걸쳐 작업자와 중장비를 탐지하고 그들의 작업상태를 파악하는 범용적 방법론을 제안하였다(Kim et al., 2024). 반면, 컴퓨터비전 AI는 건설현장 안전 모니터링에도 적용되었다. 예를 들어, 작업자가 안전모 등의 개인보호구를 잘 착용하고 있는지를 자동으로 인식하고(Ding et al., 2022), 작업자-중장비 간 물리적 거리, 움직이는 방향 및 속도를 추정하여 충돌 가능성을 추론하기도 했다(Ding and Luo, 2024).

최근에는 현장 이미지로부터 보다 세밀하고 깊이 있는 정보를 추출하려는 연구가 진행되고 있다. 그 예로, Kim and Chi(2022)은 2차원 건설현장 이미지에 나타난 물체들과 그들의 관계를 그래프 형태로 나타내고 분류하는, 그래프신경망 기법을 제안하였으며, Yan et al.(2020)의 연구에서는 건설물체 간의 3차원 관계까지 파악할 수 있는 방안을 제시하였다. 또 다른 연구에서는 이미지 분석이 어려운 상황에서

컴퓨터비전 AI의 정확도를 향상하기도 했다. 특히 분석오류의 주요 원인인 중첩상황에서 건설 작업자 또는 중장비를 어떻게 탐지할 것인지(Wang et al., 2024), 매우 먼 거리에서 촬영된 작은 크기의 물체를 인식할 것인지(Park et al., 2023) 등에 대한 연구가 이루어졌다.

이러한 성과에도 불구하고, 컴퓨터비전 AI의 성능이 학습 데이터셋에 굉장히 의존하고(염준호, 2023; 권영주·문성호, 2023) 이를 위해 너무 많은 시간과 노력이 소요되기 때문에 실제 건설현장에서 충분히 활용되지 못하고 있는 실정이다.

2.2 기존의 학습 이미지 데이터셋 구축 방법

실제 현장에서 활용 가능한 컴퓨터비전 AI를 개발하기 위해 학습 이미지 데이터셋을 어떻게 효과적, 효율적으로 구축할 것인지에 대한 연구가 이루어져 왔다. 구체적으로 다음과 같은 2가지의 대표적인 방법들이 있다.

2.2.1 합성 이미지 생성 방법

일부 연구자들은 가상의 건설현장을 모사할 수 있는 디지털 환경에서 추출된 합성 이미지에 주목했다. 디지털 환경에서는 어떤 종류의 상황도 손쉽게 재연할 수 있고 무수히 많은 양의 이미지 데이터를 뽑아낼 수 있어 그 잠재력이 매우 높게 평가되었다. 예를 들어 Braun and Borrmann(2019)은 3차원 건물정보모델링을 활용하여 기둥, 빔 등 건축물의 구성요소에 대한 합성 이미지를 생성하고 이를 활용하여 컴퓨터비전 AI 모델을 학습하는 데 성공하였다. Lee et al.(2022)은 굴삭기 탐지를 위해 이에 대한 합성 이미지를 생성하였고, Torres Calderon et al.(2021)의 연구에서는 3차원 굴삭기 모델의 작업을 물리적으로 재현하고 굴삭기의 작업유형을 판단하는 모델을 구축하였다. 이와 유사하게, Assadzadeh et al.(2022)와 Mahmood et al.(2022) 연구에서는 굴삭

기의 2차원/3차원 자세를 추정하는 데 맞춤 개발하였다. 해당 연구들이 건축물 구성요소, 중장비에 초점을 맞추었다면 Kim et al.(2023)은 건설 작업자에 대한 가상 이미지를 합성하고 이를 활용하여 컴퓨터비전 AI 모델을 훈련시켰다. 특히, 실제 건설현장에서 수집된 데이터셋과 가상의 건설현장에서 생성된 합성 이미지를 동시에 활용하여, 그 효과를 검증하였다. 또한, Ding and Luo(2023)은 작업자와 중장비가 동시에 나타나는 현장을 모사하여 이에 대한 대규모의 합성 데이터셋을 구축하였다.

이처럼 합성 이미지 데이터셋의 잠재력이 확인되었으나 합성 이미지는 본질적으로 실제 건설현장 이미지와는 그 시각적 특성이 다르기 때문에 컴퓨터비전 AI 모델의 성능을 향상하는 데 한계가 있다(Kim et al., 2023).

2.2.2 실제 건설현장 이미지 수집 및 벤치마크 데이터셋 구축

많은 연구자들은 실제 건설현장에서 다량의 이미지 데이터를 직접 수집하고 라벨링하여 대규모의 건설현장 벤치마크 이미지 데이터셋을 구축하려는 노력을 진행하였다. 예를 들어, Xiao and Kang(2021)은 굴삭기, 덤프트럭, 컴팩터, 도저, 그레이더, 타워크레인 등 총 10개 유형의 건설장비에 대한 이미지를 10,000장 수집하였고, Duan et al.(2022)에서는 작업자, 안전모, 철근 등에 대한 이미지를 약 20,000장 촬영하여 실제 이미지셋을 만들었다. 이와 유사하게, Yan et al.(2023)도 작업자, 중장비에 대한 실제 이미지를 50,000장 수집하여 공개하였으며, Xuehui et al.(2021)은 174여개의 서로 다른 건설현장에서 작업자, 굴삭기, 덤프트럭, 믹서트럭 등에 대한 41,000여장의 실제 이미지를 촬영하였다. 이 같은 데이터셋이 대부분 작업자, 중장비, 자재, 도구 등의 물체를 탐지하는 데만 초점을 맞추었기 때문에 다른 연구에서는 작업자의 불안정한 행동 인식을 위한 이미지셋

을 구축하기도 했다. Yang et al.(2023)는 총 7가지의 불안정한 행동에 대한 1,595개의 비디오를 촬영하여 공개하였다. 다른 한편으로는, 최근 친환경에 대한 사회적 수요 증가에 따라 일부 연구는 건설현장에서 발생하는 먼지들에 대한 이미지 데이터를 모으고 라벨링을 수행하기도 했다(Xiong and Tang, 2021; Wang et al., 2023).

2.3 지식격차: 학습 이미지 양과 특화성 간의 복잡한 관계 및 AI 성능에 미치는 영향

앞에서 서술한 것처럼 학습 이미지 데이터셋 구축에 관한 많은 노력이 이루어졌으나 대부분 “학습 이미지가 많으면 많을수록 AI 성능이 증가할 것이다”라는 가설에 기반하고 있다. 하지만 주변환경이 시간에 따라 끊임없이 변화하는 건설현장에서는 학습 데이터셋의 크기뿐만 아니라, 그 데이터가 해당 공종에 얼마나 특화되어있는지에 따라 AI 성능이 매우 달라질 수 있다. 최근 연구에서도 실제 컴퓨터비전 AI를 적용하는 현장과 다른 특성을 지닌 이미지를 너무 많이 활용하면, AI 모델의 성능이 오히려 감소한다는 사실이 발견되었다(Kim et al., 2023). 이 발견은 많은 양의 이미지를 확보하는 것도 중요하지만 학습 데이터셋의 분포가 시험 이미지셋의 분포와 유사하도록 만들어야 한다는 것을 시사하며, 더 나아가 시시각각 변화하는 건설현장에서는 “각 공종에 특화된 학습 이미지셋을 구축하면 데이터 수가 다소 부족하더라도 더 높은 성능의 AI 모델을 구축할 수 있다”라는 기존과 대립하는 가설을 세울 수 있다. 예를 들어, 일정량 이상의 학습 데이터를 확보하면 데이터셋 크기 자체보다는 실제 건설현장에 진행되고 있는 공종에 얼마나 특화되어있는지가 더 중요해질 것이다. 하지만 건설분야에서 학습 이미지셋의 크기와 특화성(공종별 특화), 그리고 컴퓨터비전 AI 성능 간의 복잡한 관계에 대한 연구가 부족한 실정이며, 이에 따라 어떤 특성의 학습 이미지를 얼마나 많이

수집해야 공종별 최고 성능의 컴퓨터비전 AI 모델을 개발할 수 있는지 의문으로 남아있다.

3. 연구방법 및 실험설정

위 지식격차를 해소하기 위해서 본 연구는 학습 이미지셋의 크기-특화성-컴퓨터비전 AI 성능 간의 관계를 정량적으로 분석하는 실험을 다음과 같이 2단계로 수행하며, 보다 자세한 내용은 아래와 같다.

3.1 건설현장 전 공종 범용 및 공종별 특화 이미지 데이터셋 구축

건설현장 전 공종에 대한 이미지로 구성된 범용 데이터셋(기존 방식)과 단일 공종으로만 구성된 공종별 특화 데이터셋(제안 방식)을 준비하였다. 이를 위해 실제 건설현장 벤치마크 이미지 데이터셋을 확보한 뒤 토공-기초-골조-마감공종별로 이미지를 직접 분류하였다. 구체적으로, 174개의 실제 건설현장에서 구축되었고 공개적으로 사용 가능한 Moving Objects in Construction Sites(MOCS) 데이터셋을 채택하였다(Xuehui et al., 2021). 그 다음, 연구진이 이미지에 직접 주석을 달고 공종을 분류하였으며, 4가지 공종에 해당하지 않거나 분류하기 애매한 이미지는 포함시키지 않았다. 더불어, 다른 공종 대비 마감공사에 대한 이미지가 현저히 부족했기 때문에 Google 등 인터넷에서 1,000장을 추가로 수집하였다.

위와 같은 과정으로 토공-기초-골조-마감 공종에 대한 이미지를 모두 확보한 다음, 먼저 전 공종 범용(학습)데이터셋을 구축하였다. 특히 특정 공종에 대한 편향성을 줄이고 모든 공종에 대한 대표성과 범용성을 최대한 확보하기 위해서 공종별 이미지 수가 모두 동일하도록 무작위 선별하였고, 데이터셋 크기는 작업자 수 기준 2,000명부터 4,000명, 6,000명, 8,000명, 10,000명으로 변경해가며 총 5개의 범용 데이터셋을 구축하였다. 예를 들어, 10,000명의 데

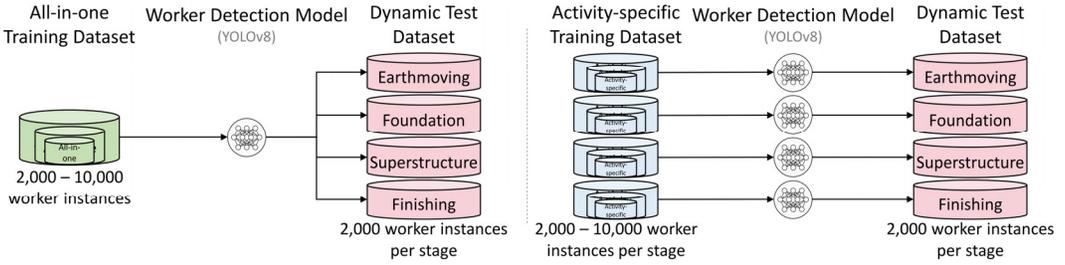


Fig. 1. Training and Test Datasets used for Training All-In-One and Activity-Specific Worker Detection Models

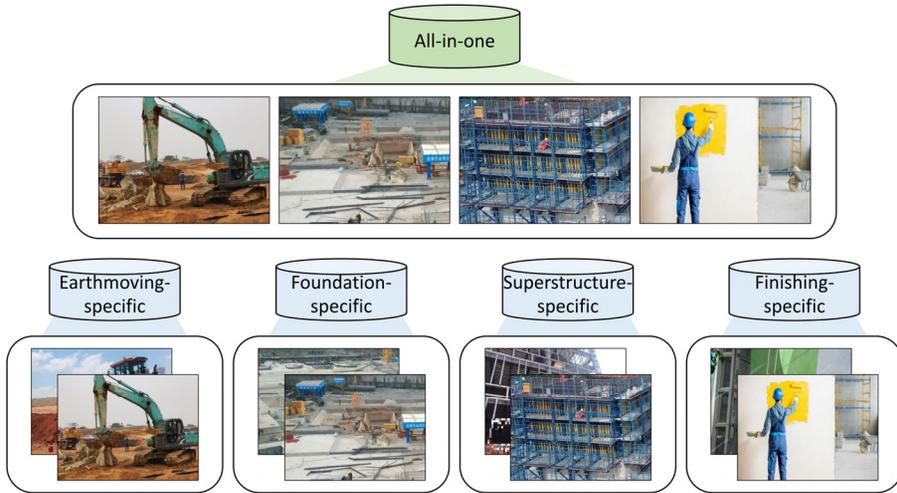


Fig. 2. Examples of Construction Worker Images included in the All-In-One and Activity-Specific Datasets

Source: MOCS Dataset by Xuehui et al. (2021)

이터셋에는 2,500명에 대한 토공, 기초, 골조, 마감 공종 이미지가 각각 포함되었다.

범용 데이터셋과 유사한 방식으로 공종별 특화(학습)데이터셋을 구축하였으며, 다만 하나의 데이터셋에는 단일 공종에 대한 이미지만 포함되었다. 예를 들어, 마감공종에 특화된 데이터셋은 마감공종에 대한 이미지로만 구성되었다. 이를 통해 총 20개의 공종별 특화 데이터셋을 준비하였다(5가지 크기의 데이터셋 크기×4가지 공종). 마지막으로, 시간에 따라 변화하는 건설현장에서의 AI 성능을 보다 면밀히 평가하기 위해서 공종별 특화 시험 데이터셋을 추가로 준비하였다. 즉, 토공-기초-골조-마감 각 공종에 대한 총 4가지의 시험 데이터셋을 구축하였으

며, 하나의 데이터셋은 2,000명의 작업자에 대한 이미지로 구성되었다.

즉, Fig. 1과 같이 전 공종 범용 및 공종별 특화 학습 데이터셋의 경우 2,000명, 4,000명, 6,000명, 8,000명, 10,000명 총 5개 크기의 작업자 이미지로 구성되었으며, 이는 다시 모델 학습과 검증을 위해 8:2로 구분되어 사용되었다. 반면에, 시시각각 변화하는 건설현장의 상황을 모사하는 시험 데이터셋은 토공 작업자 2,000명, 기초 작업자 2,000명, 골조 작업자 2,000명, 마감 작업자 2,000명 이미지로 구축되었다. 데이터셋에 포함된 이미지의 예시는 Fig. 2에서 볼 수 있다.

3.2 작업자 인식 모델 학습 및 성능 평가

앞서 구축한 학습 이미지셋의 효용성을 비교·분석하기 위해 작업자 인식 모델을 학습하고 그 성능을 평가하였다. 구체적으로, 전 공종 범용 및 공종별 특화 학습 이미지셋을 사용하여 8번째 버전의 You Only Look Once Nano(YOLOv8n) 기반의 모델을 구축하였다. YOLO는 건설분야에서 가장 많이 활용된 모델 중 하나이고 좋은 성능을 보였을 뿐만 아니라, 해당 모델의 작동 메커니즘에 대한 지식이 축적되어 왔기 때문에 연구결과의 신뢰성과 해석 가능성을 모두 확보할 수 있었다. 이를 위해 [image size: 640×640], [epochs: 300], [batch size: 16], [optimizer: AdamW with momentum of 0.9 and weight decay of 0.0005], [learning rate: 0.01]의 하이퍼파라미터를 적용하였다.

이렇게 학습된 YOLOv8n 모델은 앞서 구축된 공종별 특화 시험 데이터셋에서 성능을 평가하였으며, 이를 위해 가장 대표적인 성능지표인 Average Precision @50(AP)을 적용하였다.

4. 결과분석 및 논의

Table 1은 전 공종 범용 및 공종별 특화 데이터셋에 포함된 학습 이미지 개수를 변경하며 학습된 모델의 공종별 성능을 보여준다. 우선 모든 공종에서 동일한 양의 학습 이미지가 사용된 경우 특화 데이터셋이 범용 데이터셋보다 평균 3.8% 이상의 높은 성능을 나타냈다. 이 결과를 통해 학습 데이터셋이 시험 데이터셋과 유사한 특성을 가질수록 AI 모델이 더 잘 작동한다는 점을 다시 한번 확인할 수 있다. 특히, 10,000명의 작업자에 대한 학습 이미지가 사용된 경우 토공 이미지로만 구성된 특화 데이터셋이 전 공종 범용 데이터셋보다 AI 성능을 5.8%나 향상할 수 있었다는 점은 학습 데이터의 특화성이 미치는 영향이 상당히 중요하다는 것을 알 수 있다.

상대적으로 작은 크기의 공종별 특화 데이터셋을 사용하더라도 대규모의 범용 데이터셋과 AI 모델 성능이 유사하거나 더 높았다. Table 1을 살펴보면 모든 공종에 대한 10,000명의 작업자 이미지로 범용 데이터셋을 구성하는 것보다, 각 공종에 특화된 6,000명 작업자 이미지를 사용할 때 더 높은 성능을

Table 1. Model Performance by Dataset's Size and Specificity over the Entire Project Life Cycle

Dataset Size (# of Workers)	Dataset Specificity	Earthmoving	Foundation	Super-Structure	Finishing	Average
2,000	All-in-one	60.30%	66.70%	60.70%	86.10%	68.45%
	Activity-specific	61.90%	68.30%	62.40%	91.70%	71.08%
4,000	All-in-one	62.30%	70.80%	65.50%	89.80%	72.10%
	Activity-specific	65.50%	73.00%	68.30%	94.30%	75.28%
6,000	All-in-one	63.90%	71.90%	67.10%	91.30%	73.55%
	Activity-specific	68.50%	75.90%	71.40%	95.80%	77.90%
8,000	All-in-one	65.50%	72.50%	68.50%	93.40%	74.98%
	Activity-specific	70.40%	77.50%	74.10%	96.20%	79.55%
10,000	All-in-one	67.70%	74.40%	71.10%	94.10%	76.83%
	Activity-specific	73.50%	78.70%	74.60%	96.80%	80.90%

확보할 수 있었다. 더하여, 4,000명 작업자에 대한 공종별 특화 이미지를 사용하면 대규모의 전 공종 범용 이미지셋 대비 단 1.6% 성능 감소가 있었다. 이 결과는 각 공종에 특화된 학습 이미지를 수집하면 필요한 데이터의 양을 40~60% 가량 줄이면서도 AI 모델의 성능을 유지하거나 더 향상할 수 있음을 의미한다. 이에 따라 공종별 특화 데이터셋을 구축할 때는 전 공종 범용 데이터셋 대비 약 50%의 이미지로만 구성하여도 더 높거나 유사한 성능을 달성할 수 있을 것이다. 더불어, 본 연구에서 사용된 MOCS 데이터셋은 Fig. 2와 같이 총 174개의 서로 다른 건설현장에서 충분히 다양하고 어려운 분석조건에서 수집되었다는 점을 고려하면, 건설분야에서 상당히 일반화가 가능하고 대표성이 있는 발견이라고 볼 수 있다.

데이터셋의 유형과 관계없이 학습 이미지가 많을수록 AI 모델 성능이 증가하는 경향이 관찰되었다. 이는 데이터셋 크기-AI 성능 간 관계에 대한 기존의 지식과 일치하지만, 학습 데이터셋 크기가 클수록 특화 데이터셋의 성능 우위가 더 두드러졌다는 점이 주목할 만하다. 구체적으로, 2,000~4,000명 작업자 이미지를 수집한 경우 공종별 특화 데이터셋이 범용 데이터셋 대비 평균 2.6~3.2% 높은 성능을 보인 반면, 작업자 수가 6,000~10,000명이 되면 그 성능 차이가 4.1~4.6%로 증가하였다(Table 1). 이러한 결과는 학습 데이터가 많을수록 AI 성능이 향상되지만, 일정량 이상의 데이터가 확보되면 데이터셋 크기보다는 특화성의 영향이 더 결정적이라는 것을 시사한다.

공종에 따른 모델의 성능 변화를 살펴보면 범용 데이터셋이 특화 데이터셋보다 더 안정적이라는 것을 알 수 있으며, 이는 범용 데이터셋에는 모든 공종에 대한 이미지가 포함되었기 때문이다. 즉, 다양한 공종에 대한 이미지를 수집하면 각 공종에서 최고의 성능을 확보할 수는 없으나 토공-기초-골조-마감 전

생애주기에 걸쳐 가장 강건한 모델을 학습시킬 수 있음을 시사한다. 그러나 실제 현장에서는 미래에 수행될 공종에 대한 이미지를 수집할 수 없다는 제약조건을 유의해야 한다. 예를 들어, 토공단계에서는 마감공종에 대한 학습 이미지를 수집할 수 없기 때문에 토공-기초-골조-마감 전 공종에 걸쳐 학습 이미지를 단계적으로 업데이트해야 할 것이다. 즉, 실제 현장에서는 학습 이미지의 개수-다양성-특화성을 모두 최적화해야만 최고 성능을 확보하면서 건설사업의 전 생애주기에서 잘 작동하는 AI 모델을 구축할 수 있을 것이다.

위와 같이 흥미로운 연구결과에도 불구하고, 일부 한계점과 향후 연구가 남아있다. 예를 들어, 실제 건설현장에서는 2가지 이상의 공종이 동시에 수행되는 경우가 있다(예: 토공-기초 작업을 동시에 수행하는 경우). 그러나 본 연구에서 사용된 시험 데이터셋은 1가지 공종에 대한 이미지로만 구성되었기 때문에 해당 상황에 대한 데이터셋을 추가 구축하여 모델 성능을 비교-분석할 필요가 있다. 더불어, 연구진은 컴퓨터비전 AI의 필수기능인 작업자 인식에만 초점을 맞추었으나 향후 연구에서는 증장비, 자재 등을 탐지하거나, 행동인식, 자세추정, 단위분할 등 다른 유형의 컴퓨터비전 AI 문제를 탐구하는 것도 중요할 것이다. 이러한 노력들을 통해 학습 이미지셋의 특성(크기, 다양성, 특화성)이 컴퓨터비전 AI 성능에 미치는 복잡한 영향을 보다 잘 이해할 수 있을 것으로 기대된다.

5. 결론

본 연구는 학습 이미지셋의 크기와 특화성이 컴퓨터비전 AI 모델의 성능에 미치는 영향관계를 정량화-분석하였다. 이를 위해 토공-기초-골조-마감에 이르는 전 공종 범용 이미지 데이터셋과 공종별 특화 이미지셋을 구축하고 컴퓨터비전 AI 모델을 개발한

뒤, 이들의 성능을 공종별로 비교·분석하였다. 실험 결과는 공종별 특화 데이터셋을 활용할 때 단일 공종에서 가장 높은 성능의 AI 모델을 학습할 수 있었고, 더 나아가 범용 데이터셋 대비 40~60% 정도 적은 양의 학습 이미지로도 AI 성능이 유사하거나 더 높은 것으로 나타났다. 이 연구결과는 “학습 이미지 데이터 수가 많을수록 AI 성능이 증가할 것이다”라는 기존 가설보다는 “각 공종에 특화된 학습 이미지 데이터셋을 구축하면 데이터가 다소 적더라도 더 높은 성능의 AI 모델을 구축할 수 있다”라는 가설을 뒷받침하는 근거가 될 수 있다. 반면, 시간(공종)에 따른 AI 모델의 성능 변화를 살펴보면 전 공종 범용 데이터셋이 더 안정적이고 강건한 성능을 보였다. 이 결과는 건설사업의 전 생애주기 걸쳐 잘 작동하는 AI 모델을 개발하기 위해서는 기존 공종(예: 토공)에서 수집한 이미지를 다음 단계(예: 기초)에서 잘 활용하면 보다 효율적인 데이터셋 구축이 가능하다는 것을 의미한다. 이러한 연구발견들은 보다 강력한 컴퓨터 비전 AI를 학습하는 데 큰 기여를 할 것이다.

참고문헌

- 권영주·문성호(2023), “드론 촬영 이미지 데이터를 기반으로 한 도로 균열 탐지 딥러닝 모델 개발”, 『LHI Journal』, 14(2): 125~135.
- 염준호(2023), “무감독 SVM 분류 기법을 통한 드론 영상 경계 박스 내 차량 자동 추출 연구”, 『LHI Journal』, 14(4): 95~102.
- Assadzadeh, A., M. Arashpour, I. Brilakis, T. Ngo and E. Konstantinou (2022), “Vision-based Excavator Pose Estimation Using Synthetically Generated Datasets with Domain Randomization”, *Automation in Construction*, 134: 104089.
- Braun, A. and A. Borrmann (2019), “Combining Inverse Photogrammetry and BIM for Automated Labeling of Construction Site Images for Machine Learning”, *Automation in Construction*, 106: 102879.
- Ding, Y. and X. Luo (2024), “A Virtual Construction Vehicles and Workers Dataset with Three-Dimensional Annotations”, *Engineering Applications of Artificial Intelligence*, 133: 107964.
- Ding, Y., Liu, M., and Luo, X. (2022), “Safety Compliance Checking of Construction Behaviors Using Visual Question Answering”, *Automation in Construction*, 144: 104580.
- Duan, R., H. Deng, M. Tian, Y. Deng and J. Lin (2022), “SODA: A Large-scale Open Site Object Detection Dataset for Deep Learning in Construction”, *Automation in Construction*, 142: 104499.
- Kim, J., and S. Chi (2019), “Action Recognition of Earthmoving Excavators Based on Sequential Pattern Analysis of Visual Features and Operation Cycles”, *Automation in Construction*, 104: 255~264.
- Kim, J., and S. Chi (2022), “Graph Neural Network-Based Propagation Effects Modeling for Detecting Visual Relationships among Construction Resources”, *Automation in Construction*, 141: 104443.
- Kim, J., D. Kim, S. Lee and S. Chi (2023), “Hybrid DNN Training Using both Synthetic and Real Construction Images to Overcome Training Data Shortage”, *Automation in Construction*, 149: 104771.
- Kim, J., J. Hwang, I. Jeong, S. Chi, J. O. Seo and J. Kim (2024), “Generalized Vision-Based Framework for Construction Productivity Analysis Using a Standard Classification System”, *Automation in Construction*, 165: 105504.
- Kim, J., S. Chi and J. Seo (2018), “Interaction Analysis for Vision-Based Activity Identification of Earthmoving Excavators And Dump Trucks”, *Automation in Construction*, 87: 297~308.
- Lee, J. G., J. Hwang, S. Chi and J. Seo (2022), “Synthetic Image Dataset Development for Vision-Based Construction Equipment Detection”, *Journal of Computing in Civil Engineering*, 36(5): 04022020.
- Luo, X., H. Li, Y. Yu, C. Zhou, and D. Cao (2020), “Combining Deep Features and Activity Context to Improve Recognition of Activities of Workers in Groups”, *Computer-Aided Civil and Infrastructure Engineering*, 35(9): 965~978.
- Mahmood, B., S. Han and J. Seo (2022), “Implementation Experiments on Convolutional Neural

- Network Training Using Synthetic Images for 3D Pose Estimation of an Excavator on Real Images”, *Automation in Construction*, 133: 103996.
16. Park, M., D. Q. Tran, J. Bak and S. Park (2023), “Small and Overlapping Worker Detection at Construction Sites”, *Automation in Construction*, 151: 104856.
 17. Roberts, D., W. T. Calderon, S. Tang, and M. Golparvar-fard (2020), “Vision-Based Construction Worker Activity Analysis Informed by Body Posture”, *Journal of Computing in Civil Engineering*, 34(4): 04020017.
 18. Torres Calderon, W., D. Roberts and M. Golparvar-Fard (2021), “Synthesizing Pose Sequences from 3D Assets for Vision-Based Activity Analysis”, *Journal of Computing in Civil Engineering*, 35(1): 04020052.
 19. Wang, M., G. Yao, Y. Yang, Y. Sun, M. Yan and R. Deng (2023), “Deep Learning-Based Object Detection for Visible Dust and Prevention Measures on Construction Sites”, *Developments in the Built Environment*, 16: 100245.
 20. Wang, Q., H. Liu, W. Peng, C. Tian, and C. Li (2024), “A Vision-Based Approach for Detecting Occluded Objects in Construction Sites”, *Neural Computing and Applications*, 36(18): 10825~10837.
 21. Xiao, B. and S.-H. Kang (2021), “Development of an Image Data Set of Construction Machines for Deep Learning Object Detection”, *Journal of Computing in Civil Engineering*, 32(2): 05020005.
 22. Xiong, R. and P. Tang (2021), “Machine Learning Using Synthetic Images for Detecting Dust Emissions on Construction Sites”, *Smart and Sustainable Built Environment*, 10(3): 487~503.
 23. Xuehui, A., Z. Li, L. Zuguang, W. Chengzhi, L. Pengfei, and L. Zhiwei (2021), “Dataset and Benchmark for Detecting Moving Objects in Construction Sites”, *Automation in Construction*, 122: 103482.
 24. Yan, X., H. Zhang, and H. Li (2020), “Computer Vision-Based Recognition of 3D Spatial Relationship between Moving Objects for Monitoring Struck-By Accidents”, *Computer-Aided Civil and Infrastructure Engineering*, 35(9): 1023~1038.
 25. Yan, X., H. Zhang, Y. Wu, C. Lin and S. Liu (2023), “Construction Instance Segmentation (CIS) Dataset for Deep Learning-Based Computer Vision”, *Automation in Construction*, 156: 105083.
 26. Yang, M., C. Wu, Y. Guo, R. Jiang, F. Zhou, J. Zhang, and Z. Yang (2023), “Transformer-Based Deep Learning Model and Video Dataset for Unsafe Action Identification in Construction Projects”, *Automation in Construction*, 146: 104703.
 27. Ding, Y. and X. Luo (2023.5.29), “Monocular 2D Camera-based Proximity Monitoring for Human-Machine Collision Warning on Construction Sites”, arXiv, <https://arxiv.org/abs/2305.17931>.

요 약

컴퓨터비전 AI는 건설산업의 디지털 전환과 로봇 자동화의 중요한 요소로, 기존에는 학습 이미지 데이터가 많으면 많을수록 모델의 성능이 더 높아진다는 가설을 기반으로 활발히 연구되어왔다. 그러나 시간에 따라 토공-기초-골조-마감작업이 진행되는 건설현장에서는 각 공종에 특화된 이미지를 사용하는 것이 데이터셋의 크기를 확장하는 것보다 더 중요하다는 대안 가설을 세울 수 있다. 이러한 배경에서 본 연구는 학습 이미지 데이터셋의 크기와 특화성(각 공종에 특화된 정도)이 컴퓨터비전 AI 성능에 미치는 영향을 조사한다. 구체적으로, 학습 이미지 개수를 변경해 가며 전 공종 범용 이미지셋과 공종별 특화 데이터셋을 각각 준비하고, 영상 기반 작업자 탐지 모델을 학습시킨 뒤 공종별 시험 이미지셋을 활용하여 그 성능을 평가·비교분석한다. 또한, 공종별 컴퓨터비전 AI의 최고 성능과 전 생애주기에서의 모델 성능 변화, 필요한 학습 데이터 개수를 종합적으로 분석하여, 건설현장 컴퓨터비전 AI 개발을 위한 최적의 학습 데이터셋 크기와 특화성을 결정한다. 연구결과는 앞서 언급한 두 가지 가설을 과학적으로 검증할 수 있는 기초자료가 될 것이며, 더 나아가 계속해서 변화하는 건설현장에 활용할 AI 모델을 위해 학습 이미지 데이터셋을 어떻게 구축하고 업데이트해야 하는지에 대한 이론적 기반을 마련할 것이다.

주제어: 건설현장, 컴퓨터비전, AI, 학습 데이터셋, 공종별 특화 이미지