

## 4차 산업혁명시대의 디지털 경공업

허 석 재<sup>1</sup>·나 승 욱<sup>2</sup>·한 세 희<sup>3</sup>·신 윤 수<sup>4</sup>·이 상 현<sup>5†</sup>

<sup>1</sup>단국대학교 리모델링연구소 연구교수, <sup>2</sup>단국대학교 건축학부 초빙교수, <sup>3</sup>단국대학교 건축학부 겸임교수,  
<sup>4</sup>단국대학교 건축학부 박사후과정, <sup>5</sup>단국대학교 건축학부 교수

### Flip Side of Artificial Intelligence Technologies: New Labor-Intensive Industry of the 21st Century

Seokjae Heo<sup>1</sup>, Seunguk Na<sup>2</sup>, Sehee Han<sup>3</sup>, Yoonsoo Shin<sup>4</sup> and Sanghyun Lee<sup>5†</sup>

<sup>1</sup>Research Assistant Professor, Remodeling Research Center, Dankook University, Yongin, 16890, Korea

<sup>2</sup>Visiting Professor, Department of Architectural Engineering, Dankook University, Yongin, 16890, Korea

<sup>3</sup>Adjunct Professor, Department of Architectural Engineering, Dankook University, Yongin, 16890, Korea

<sup>4</sup>Postdoctoral Researcher, Department of Architectural Engineering, Dankook University, Yongin, 16890, Korea

<sup>5</sup>Professor, Department of Architectural Engineering, Dankook University, Yongin, 16890, Korea

#### Abstract

The paper acknowledges that many human resources are needed on the research and development (R&D) process of artificial intelligence (AI), and discusses on factors to consider on the current method of development. Enfin, in order to enhance efficiency of AI development, it seems possible through labour division of a few managers and numerous ordinary workers as a type of light industry. Thus, the research team names the development process of AI, which maximizes production efficiency by handling digital resources named 'data' with mechanical equipment called 'computer', as digital light industry of fourth industrial era. As experienced during the previous Industrial Revolution, if human resources are efficiently distributed and utilized, digital light industry would be able to expect progress no less than the second Industrial Revolution, and human resources development for this is considered urgent.

**Keywords** : digital light industry, fourth industrial era, AI, human resources development, work index, data waste

#### 1. 서론

대한민국의 기업가들은 반세기 동안 농업 중심의 국가를 경공업, 중공업 국가로 3차산업혁명 시기에는 첨단 전자 산업 국가로 도약시켰다. 2000년 이후에는 세계 정상에 오른 기업도 속속 등장하였다. 지금 그들은 4차 산업혁명의 글로벌 선도기업으로 거듭나기 위해 AI 연구센터를 설립하고 공격적으로 전문인재 육성하고 있다. 이제는 어느 정도 성과를 보이는 기업체도 나타나고 있다. 특히 건설, 교통 분야에서는 Naver Labs가 발표한 인공지능 실내 맵핑기술과 측위기술이 세계 정상급으로 인정받고 있으며(Revaud *et al.*, 2019), 카카오브레인은 이미지에 대한 레이블 없이도 이미지를 스스로 학습하는 방법에 대한 세계 최고 수준의 성능을 가지는 기법을 개발하였다(Kim

*et al.*, 2021; Roh *et al.*, 2021). 이에 따라 전문가들은 인공지능이 전반적인 산업 분야에 곧 적용하게 될 것이라고 예상하였고(King *et al.*, 2019), 국가적으로도 2020년 하반기 '디지털 뉴딜'의 대표과제인 '데이터 댐' 사업을 본격적으로 시작하였다(NIA, 2020). 2020년부터 시작된 이 사업은 2021년 상반기 총 사업액이 2,925억원으로 84개 분야의 학습용 데이터를 수집한다. 이 사업의 취지는 학습데이터 생산 업무가 전체 개발시간의 80% 이상을 차지하여 연구개발이 지연되기 때문에 이를 국가적으로 대신하여 기업체와 연구기관이 인공지능연구에 몰입할 수 있도록 하는 것이다. 본 사업을 통해 전 산업 분야에 인공지능 혁신을 촉진하고 대규모 일자리를 창출하고 있는 긍정적인 효과가 보이고 있다. 이와 같은 국가 산업 단위의 데이터 수집을 통해 대한민국은 빠른 시점 안에 인공지능 강국으로 도약할 수

†Corresponding author:

Tel: +82-31-8005-3754; E-mail: lshyun00@dankook.ac.kr

Received August 17 2021; Revised September 23 2021;

Accepted September 24 2021

© 2021 by Computational Structural Engineering Institute of Korea

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

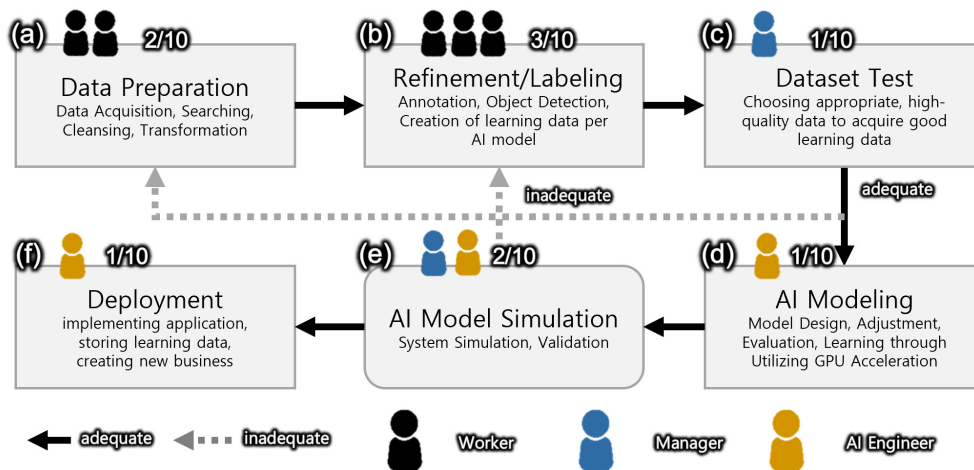


Fig. 1 The current AI R&D process in academia and in industrial field, and manpower status required for development (If 10 personnel are needed for total input, 50% of them are for data acquisition, processing, refinement and Labeling)

있을 것으로 예상된다.

인공지능 연구개발 방식은 이미지를 사용하는 방식과 시계열 데이터를 활용하는 방식으로 구분할 수 있다. 이 두 가지 모두 일반적인 수집, 정제, 라벨링 과정을 거치며 Fig. 1은 개발 프로세스에서 데이터 수집과 가공, 정제와 검증 부분에 중점을 맞춘 계획도이다.

일반적으로 원천 데이터를 획득하고 난 후 데이터의 크기, 형태를 조절하는 정제과정(Refinement)과 어노테이션(Annotation), 세그멘테이션(Segmentation)과 같은 라벨링(Labeling) 작업, 학습하려는 AI 모델이 맞추어 학습데이터(Dataset)을 생성하는 과정이 필요하다. 그 다음 정제된 데이터를 인공지능의 뇌라고 할 수 있는 가중치 모델에 재분류한 후 마지막으로 최적의 결과를 얻기 위해 품질평가를 실시하고 사용여부를 정하게 된다 (Dosovitskiy *et al.*, 2020; Felzenszwalb *et al.*, 2008; Girshick *et al.*, 2014; Girshick, 2015; He *et al.*, 2016; He *et al.*, 2017; Huang *et al.*, 2017; Krizhevsky *et al.*, 2012; Lin *et al.*, 2014; Lowe, 1999; Redmon *et al.*, 2016; Ren *et al.*, 2015; Simonyan and Andrew, 2014; Tan and Quoc, 2019; Vaswani *et al.*, 2017).

대부분의 이전 논문에서는 개발과정을 원시데이터 획득-전처리가공-데이터정제-데이터라벨링-데이터셋 구성으로 세분화하여 구분하였다. 이 과정은 결국 전처리와 데이터셋 구성을 제외한 데이터 획득/정제/라벨링 단계로 묶어 간소화할 수 있다. 그 이유는 학계가 아닌 산업계의 인공지능 연구개발은 비용절감과 시간단축을 위해 데이터 획득과 동시에 가공을 하거나 정제과정에서 라벨링 작업까지 수행하기 때문이다. 또한 학습용 모델을 자체적으로 개발하기보다는 이미 개발된 모델을 사업목적에 맞추어 선정하고 합한 이미지, 동영상 등의 데이터를 모은 후 빠르게 정제와 검증과정을 거치게 되기 때문에 연구기관에서 수행하는 세분화된 프로세스를 따르지 않는다 (Kim *et al.*, 2019; Lee *et al.*, 2007). 여기서, 말하는 검증은 일반

적으로 인공지능 연구에서 말하는 검증집합(Validation set)을 구성하는 것이 아니라 정제과정이 정상적으로 이루어졌는지 인력 또는 자동화도구를 사용하여 확인하는 과정을 말한다.

위 과정을 따르면 인공지능 연구개발이 쉽고 빠르게 이루어질 것 같으나 인공지능 연구에는 고려해 할 사항이 2가지 존재한다. 첫 번째는 학습단계에서 많은 고급 연구인력과 장비가 필요하다는 것이다. 학습모델의 성능을 확인하기 위해서는 수많은 변수연구가 수행되어야 하며, 모델의 구조를 이해하고 변경할 수 있는 고급 연구자와 학습시간 단축을 위한 장비, 라벨링을 수정할 수 있는 인력이 갖추어져야 한다. 현 대한민국에서 인공지능 구조를 제대로 알고 설계할 수 있는 고급인력이 충분하지 않으며 가용가능한 인원 또한 특정분야에만 치중되어 있다. 특히 건설, 토목과 같은 비주류 분야에서는 연구개발 수요는 계속 발생하고 있으나 인력부족으로 연구개발이 더디게 진행되고 있다. 개발장비의 문제도 존재한다. 최신 인공지능 모델을 개발하기 위해서는 하드웨어, 특히 GPU의 수량이 중요하다(Baji, 2018). 여러 대의 GPU를 병렬로 사용하여 계산을 수행하는 인공지능 기법이 존재하며, 이에 따라 수량이 많으면 많을 수록 변수연구가 수월해진다. 하지만 건축, 토목공학과 같이 비주류 분야를 연구하는 연산장치는 Table 1과 같이 2020년 시점 4000~10000 \$ 내외 정도의 장비로 연구를 수행하는 것으로 보이며, 장비보유없이 구글이나 아마존에서 제공하는 서비스를 활용하기도 하였다(Na *et al.*, 2021; Heo *et al.*, 2018). 이와 같이 비주류 분야는 낮은 성능의 연산장비로 인해 연구개발이 지연되는 실정이다.

두 번째는 데이터 수집과 정제과정에서 학습단계보다 많은 일반인력이 소모된다는 것이다. 데이터 수집을 위해 웹에서 크롤링하거나 기존데이터를 재활용하는 방식에는 한계가 있으며 특수분야의 경우 학습용 데이터 수집조차 어려운 경우가 있다. 예를 들어 지진, 태풍과 같은 재난상황에 대한 피해구조

물의 데이터를 수집하기 위해서는 적절한 시기에 연구인력이 시간과 비용을 투자하여 현장에서 직접 데이터를 수집해야만 한다. 이 과정을 거쳐 다행히 학습용데이터를 수집하더라도 최신 인공지능 모델 학습에는 정제과정이 필수적이며 이 과정은 인간의 개입을 통해 학습데이터를 각각 수정해야 한다. 특히 전문분야의 지식이 필요한 데이터 정제과정은 제한된 인원으로 인해 연구진행이 늦어지기도 한다.

기업체에서는 비용절감을 위해 데이터 수집과 정제보다는 학습분야에 치중하여 연구를 진행하고 있으나 좋은 학습데이터를 구할 수 없어 성과가 미비한 실태이다. 다행히 대한민국에서는 인공지능 모델개발 연구뿐만 아니라 데이터 수집과 정제에 많은 투자를 하고 있다. 2021년 2월에는 데이터 구축, 정제 메뉴얼을 한국지능정보사회진흥원에서 발간하여 개발방식의 표준화를 시도했다(TTA, 2021a, 2021b). 다만 수집하는 데이터의 형태가 이미지와 동영상 데이터에 맞추어 있고, 개발목적은 고려하지 않은 단순한 데이터 ‘양’ 늘리기에 치중하고 있어 본래의 수집목적이 아닌 다른 용도로 사용이 어려운 실정이다. 추후 이 데이터를 유의미하게 활용하려면 다시 인력과 비용을 투입하여 원천데이터를 재정제해야 하는 문제가 있다. 따라서 본 논문에서는 현시점에서 이루어지는 이미지 기반의 인공지능 연구의 인력 의존적인 문제점과 이를 개선하고 나아가야 할 방향에 대해 논의한다.

**Table 1** Current status and cost of AI research equipment for architectural and civil engineering majors in universities in Korea(2020)

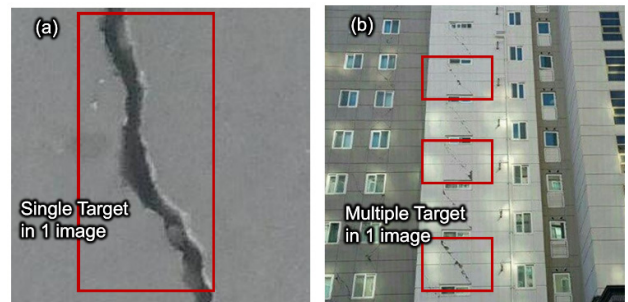
University	Major	Number of Manpower	Hardware	Cost
D	Civil Engineering	3	RTX3080 x 4	\$12,000
	Environmental Engineering	3	RTX2090 x 3	\$6,000
S	Environmental Engineering	5	RTX2060 x 4	\$4,000
H	Construction	2	Google API	\$10 / month

## 2. 데이터 수집과 정제, 품질평가에 대한 기존연구

### 2.1 인간의 시각정보를 모사하는 인공지능연구

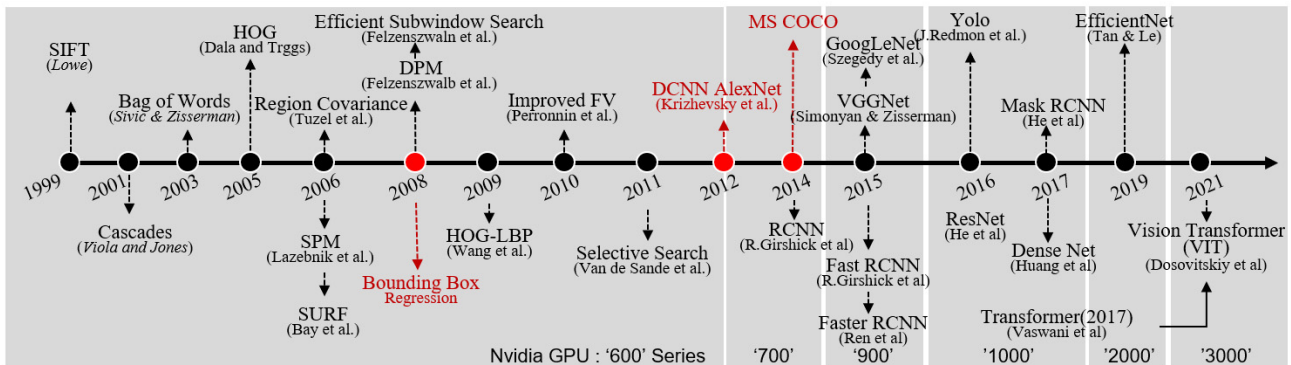
사진 이미지나 동영상의 잘려진 프레임 이미지는 인공지능 학습을 위한 좋은 자료가 된다. Fig. 2와 같이 인공지능 개발 초창기인 2008년 이전까지는 단순 이미지 분류(Image Classification) 위주의 연구개발이 이루어졌으며, 이 시기에는 이미지에 텍스트 라벨을 추가하는 수준의 라벨링이 이루어져 학습데이터를 만드는데 많은 인력이 필요하지 않았다. 이후 정확도를 향상시킬 수 있는 객체탐지(Object Detection) 기법이 개발되면서, 이미지 위에 학습에 필요한 부분을 표시해주는 추출단계(Annotation)가 필수적이 되었다(Felzenszwalb *et al.*, 2008). 초기단계의 추출방법은 이미지 분류에서 사용하였던 데이터에 테두리(Round Boxing) 등의 표시를 하는 수준이었지만 Fig. 3과 같이 한 이미지 상에서 여러 개의 물체를 구분할 수 있게 되어 효율성이 향상되었다.

하지만 Fig. 3(a)와 같이 추출단계를 고려하지 않고 수집된



\* (a) Data can also be used for object detection but it is less efficient and may lead to inaccurate results.

**Fig. 3** Difference between (a) an example of learning data collected in the era when image classification techniques were dominant and (b) learning data collected after object detection was developed



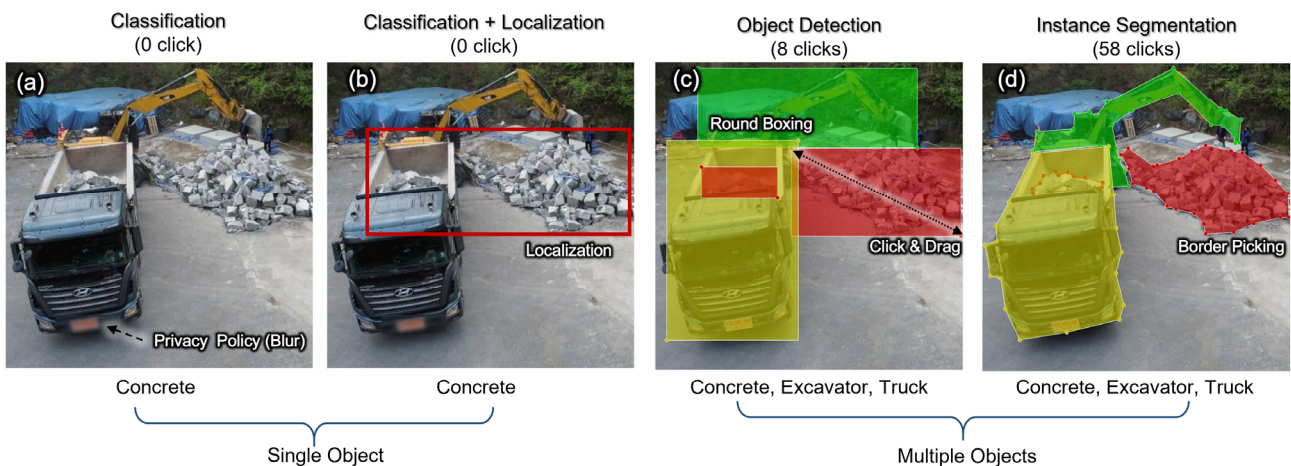
**Fig. 2** Milestones of object detection techniques and the evolution of hardware(GPU)

이전 기법의 데이터를 사용할 경우 추출방식의 장점을 살리지 못하여 인공지능의 정확도에 큰 차이를 보이지 않았다. 따라서, 객체탐지 기법을 개발하는 연구자들은 Fig. 3(b)와 같이 기법에 맞는 새로운 이미지 데이터를 수집해야만 했으며, 적용하려는 인공지능 모델에 따라 비정형 정제작업을 다시 수행해야 되는 어려움이 있었다.

2012년에는 인공지능 모델의 정확도를 높이려는 연구자들이 객체탐지 기법의 일종인 픽셀 별로 클래스를 구분하는 이미지 분석(Instance Segmentation)방식을 개발하였고 최근 연구자들은 대부분 이 기법을 사용하여 연구하고 있고 최고의 성능을 보이고 있다(Dosovitskiy *et al.*, 2020; Girshick, 2015; He *et al.*, 2016; He *et al.*, 2017; Huang *et al.*, 2017; Redmon *et al.*, 2016; Ren *et al.*, 2015; Tan and Quoc, 2019; Vaswani *et al.*, 2017). 이미지 분석기법은 이미지에서 학습에 필요한 부분을 픽셀단위로 특정하는 방식으로 라벨링을 수행하기 때문에 주로 Fig. 4(d)의 다각형 형태로 라벨링을 수행한다. 이 방식의 경우 기존 방식(Round Boxing)에 비하여 작업량이 수배로 늘어나게 된다. 이를 보완하기 위해 소위 자동라벨링 불리는 기법들이 개발되기 시작하였으며 나무, 신호등과 같은 일반사물들은 클릭이 없이도 라벨링이 되는 수준에 이르렀다(Acuna *et al.*, 2018). 하지만 오류가 많이 발생하여 사람의 보완작업이 필요하며 만일 목표로 하는 대상을 완벽하게 인식하기 위해서는 이미지를 추가로 수집하고 학습해야 할 것이다. 추가로 모아야 하는 수집량은 개발모델과 목표로 하는 정확도에 따라 달라지며 개발자마다 의도하는 숫자는 다를 수 있다. 학습데이터 수집방식은 웹서핑을 통한 온라인 수집과 직접 촬영과 같은 오프라인

수집방법으로 나뉜다(TTA, 2021a). 온라인 수집은 사람들이 자체적으로 구축한 이미지서버에 직접 웹검색을 통해 접근하여 수집하거나 크롤링 기술을 활용한 자동 웹검색 기능을 많이 활용하고 있다. 산업계에서는 저작권 관련된 문제로 인하여 오프라인 수집을 선호하며 연구개발단체에서는 개발일정과 비용을 단축하기 위해 온라인 수집과 오프라인 수집을 병용한다.

수집된 데이터는 원천데이터라고 불리며 현 단계에서는 인공지능 학습에 사용하기에 부적합한 상태이다. 원천데이터에서 학습에 필요한 데이터를 구분(Data Cleansing)한 후 학 개발하는 모델에 맞게 데이터를 수정해야만 하며 이 때 발생하는 작업을 정제(Data Refinement)라고 부른다. 정제는 기본적으로 카테고리 폴더를 생성하고 이미지를 분류하는 작업부터 이미지의 크기를 조절하거나 해상도를 변경, 개인정보보호를 위한 흐림처리, 색상을 단순화하기 위한 이진화와 같은 작업이 있으며 이후 이미지에 목표로 하는 물체를 표시하는 이미지박싱, 세그멘테이션 등의 라벨링 과정을 거친다. 각각의 작업은 적합한 프로그램 툴이 존재하며 한 번에 원하는 작업을 마칠 수 없기 때문에 아직 자동화가 이루어지지 않은 기술이라 할 수 있다. 정제에서 가장 많이 사용되는 작업은 이미지의 크기, 해상도 조정, 파일의 이름 변경이며 이 작업후에는 다시 라벨링 작업이 수행될 수 있다. 이 과정 중 인력투입이 지속적으로 발생한다. 또한 건설, 토목, 의료 등 연구분야에 따라 정제와 라벨링 작업에 전문적인 지식이 필수적인 경우도 있다. 하지만 전문교육을 받은 인력은 비싸고 수가 한정되어 있어 특수분야 인공지능 연구성과가 낮거나 지연이 되는 실정이다.



\*( ) Inside the parentheses infer the number of mouse clicks.  
 \*License plate, human faces were mosaicked or blurred for privacy.  
 \*After being able to perceive various objects on one image, the number of mouse clicks as well as refinement time increased whereas the volume of the training dataset decreased, leading to quickened learning speed, thus need comparative study on this.

Fig. 4 A typical artificial intelligence method to perceive image

## 2.2 수집데이터와 인공지능 모델의 품질관리 연구

품질관리는 두가지로 구분하여 인공지능 모델의 정확도를 높이기 위해 학습에 적합한 데이터를 추출하는 작업과 개발된 인공지능 모델의 성능평가를 위해 수행하는 작업으로 나뉜다 (TTA, 2021b). 먼저 분류모델에 대한 성능평가 지표로는 정확도(Accuracy), 오차행렬(Confusion Matrix), 정밀도(Precision), 재현률(Recall), F1 Score, ROC AUC, mAP 등이 있으며 실제 데이터에서 예측데이터가 얼마나 유사한지 판단하는 방식을 기반으로 만들어졌다. 인공지능 태동기라 할 수 있는 2012년 당시에는 기계학습 모델에만 정확도에만 치중하여 학습데이터의 품질관리가 중요하게 생각되지 않았으나 보관되는 데이터의 양이 늘어나고 정제와 라벨링과정의 중요성이 확인되면서 품질평가의 필요성이 늘어났다. 품질평가가 일반적인 업무 프로세스에 포함된다면 최소 인공지능개발 지식을 보유한 인력들이 수행 가능할 것이며 이는 인공지능 개발자들이 학습이 전단계의 데이터 품질평가 업무를 추가로 부담하게 된다는 말과 같다. 최신연구에 따르면 질 좋은 학습데이터 세트만으로도 모델의 정확도를 증진시킬 수 있다는 것이 검증되고 있는 추세로 앞으로 학습데이터 평가지표에 대한 연구와 관련 인력이 양성될 경우 해결되겠지만 단기적으로 인공지능 개발인력 부족이 문제로 떠오를 것이다.

## 3. 왜 디지털 경공업인가?

### 3.1 인력 의존적인 원천 데이터 수집방식

최근 어떠한 분야에서나 인공지능 연구가 활발히 이루어지고 있으나 학습용 데이터 수집에 대한 관심은 상대적으로 적다. 수집방법은 주로 직접촬영 등을 통한 수집방법과 웹크롤링을 통한 방법으로 나뉜다. 여기서, 많은 연구개발 기획자들이 하는 착각은 인터넷의 무한한 데이터를 활용하고 근처 주변환경에서 보이는 모든 사물을 사용할 수 있는 자료로 판단하여 데이터 수집과정이 어렵지 않고 중요하지 않다고 생각한다. 실상 전문분야의 경우 데이터 1개를 수집하기 위해 대학수준의 전공지식과 그에 준하는 검증이 필요하다. 다행히 데이터를 수집할 수 있는 기회가 오히려 특별한 제약사항이 뒤따르는데 의료분야의 경우 환부의 이미지데이터를 획득하기 위해 수많은 임상케이스와 개인의 신상정보 유출을 막기위한 노력이 필수적이다. 건설분야의 경우 데이터 수집시 개인의 사유재산을 침범하지 않도록 유의해야 하며, 이외의 분야도 이미지 상에서 지식저작권과 개인정보유출, 보안사항 유의하면서 데이터를 수집해야만 한다. 따라서, Table 2와 같이 원활한 수

Table 2 Parts to be considered for verification in data acquisition and cleansing

Parts to Verify	Processing Method
Copyright infringement	Data deletion
Personal information infringement	Blur image, delete personal information
Security facilities, public facilities	Acquire data upon prior permission
Discrimination against particular region, society and race	Set up procedure to avoid inequality
Possibility for safety related accident outbreak	Prior check on accident risks and give safety education

집이 이루어지더라도 연구개발 윤리수칙을 준수하기 위해서는 수집된 원천데이터에서 개인정보침해우려 부분을 흐림처리 하거나 저작권을 위반하는 데이터는 삭제 등이 필요하며 이는 전문인력의 수작업으로만 가능하다. 실제 상용화하기 위한 데이터라면 훨씬 더 꼼꼼한 작업이 필요하며, 2021년에 일어난 인공지능 챗봇의 개인정보보호법 위반과 같은 사례가 일어나지 않을 것이다.

직접수집기법이 아닌 웹사이트의 검색엔진을 활용한 디지털 크롤링 기술은 쉽고 간편해보이나 단점이 많아 전문개발진들은 사용을 줄이는 추세이다. 중복되거나 연관성이 없는 데이터, 다양한 포맷, 해상도 품질, 저작권 문제로 인해 수집 후 원천데이터화 하기위해 개별데이터의 검수작업이 필요하며 이에 들어가는 노력이 직접 수집하는 것만큼 많이 필요하다. 현재 인공지능 데이터를 수집하는 많은 초급작업자들은 웹 온라인상에 수많은 학습 데이터가 존재할 것이라 생각하고 웹기반의 데이터 수집을 선호하나 일정수량 이상의 데이터를 수집하면 한계에 부딪혀 결국 직접수집으로 연구방향을 선회하기도 한다. 하지만 두 방법 모두 시도하지 않을 수 없는 데이터 수집방식이며 이와 같은 시행착오로 인해 데이터 수집과정에 막대한 인력과 비용이 필요한 실정이다.

### 3.2 데이터 가공/정제/라벨링 전문인력의 필요

앞 절에서 기술한 것과 같이 이미지 데이터 가공, 정제는 일반적으로 인력을 활용한 수작업에 의해 이루어진다. 수집된 데이터의 카테고리를 분류하는 작업부터 가공, 정제단계라고 판단한다면 이 작업 또한 데이터베이스 담당자가 이미지를 하나하나 분류해야 한다. 다음 단계인 초기가공(Cleansing), 정제(Size transform, Binarization, Naming)작업은 프로그래머의 간단한 프로그램코드로 해결 가능하나 프로그램 인력을 보유하지 않았을 경우 데이터베이스관리자가 직접 개별로 작업을 수행할 수도 있다. 최신 인공지능 기법을 활용하기 위해 수행되는 라벨링은(Round Boxing, Segmentation)은 데이터의 종류에

**Table 3** Research process, human resources and level of academic knowledge required for AI research and development prior to learning

Step	Occupational Group	Level of academic knowledge
Data Acquisition	DB manager	Low level of computer programming major
	Worker	Major not required
Data Refinement	Programmer	General level of programming major
Data Labeling (Annotation)	Labeling Manager	High level of special field major
	Worker	Major not required
Refinement/Labeling Program Develop	Programmer	High level of programming major
Training data transformation	AI Engineer	High level of AI major

따라 전공지식이 필요할 수 있으며 모든 인력을 고급화한다면 비용적으로 매우 비효율적일 수 있다. 따라서, 전문지식을 보유한 관리자가 여러 명의 일반작업자를 두고 가공/정제 작업을 실시하는 방식으로 해결할 수 있다.

이 단계에서 라벨링을 위해 특별한 프로그램이 필요하기도 한다. 비영리 연구기관은 주로 ‘Labelme’(Russell *et al.*, 2008) 나 ‘VGG annotator’(Dutta *et al.*, 2016)와 같은 오픈소스를 기반으로 한 정제프로그램을 사용하나 사업을 영위하는 단체는 자체 개발한 프로그램을 사용하기도 하며 이를 위한 관리인력이 소요된다. 마지막으로 정제된 데이터를 관리하고 학습데이터화 하기 위해 인공지능 전문개발인력이 필요하며 Table 3과 같이 데이터 가공, 정제, 라벨링과정에서만 최소 5가지 지식수준의 작업인력이 필요할 것으로 판단된다.

물론 기술개발이 진행되면서 자동으로 라벨링을 도와주는 프로그램이 발표되고 있으며(Acuna *et al.*, 2018) 점점 그 활용도는 높아질 것이다. 하지만 현시점에서는 데이터 수집부터 모든 가공, 정제단계에 인간의 개입이 없을 수 없으며 인공지능 연구개발산업은 매우 인력 의존적임을 알 수 있다.

### 3.3 범용성이 떨어지는 데이터 수집/가공 방식

일반적으로 인공지능 연구개발은 한번의 데이터 정제로 이루어지지 않는다. 상용화 단계에 도달하기 위해 학습데이터는 여러번의 재가공, 정제, 라벨링하는 과정이 필연적이며 다수의 작업자 참여로 인한 불균일적인 데이터 관리가 발생하기도 한다. 재가공, 정제 원인은 크게 3가지로 나눌 수 있는데, 먼저 전문인력이라고 하더라도 개발자와 데이터관리자의 의도와 벗어나는 방식으로 라벨링을 수행하는 경우이다. 이때는 개발

목표에 맞게 일련의 정제과정을 다시 수행해야만 한다. 두번째로 학습 과정에서 정확도가 떨어지거나 분류된 모델이 원하지 않는 대상을 분류하는 경우이다. 이 경우 모델학습과정에 발생한 문제와 다르게 그 원인이 확실치 않으므로 전면적으로 전체 데이터의 재가공, 정제가 필요할 수 있어 프로젝트 일정에 큰 영향을 미친다. 원인으로 추론 가능한 것은 정제, 라벨링 작업의 정밀도(불필요한 데이터의 라벨링, 개발의도와 다른 데이터의 삽입)가 떨어지는 것과 학습 데이터 숫자의 부족일 것으로 예상할 수 있다. 마지막 세번째는 새로운 활용목적의 인공지능 네트워크 개발이다. 학습용 데이터는 한 묶음의 카테고리 구성되어 있어 별도의 카테고리를 추가, 삭제하는 것은 어렵지 않으나 새로운 목적의 인공지능 개발에 묶음으로 된 데이터를 그대로 가져다 쓰기는 어렵다. 예를 들어 건설자재라도 사용 정도에 따라 현장에서는 제품이 될 수 있고 폐기물처리장에서는 쓰레기가 될 수 있다. 자재를 인식하는 인공지능이 성공적으로 개발되었더라도, 자재관리를 위해서는 새 상품과 사용된 상품을 구별해야만 하며 이 경우 기존의 카테고리(클래스)와 데이터셋을 수정해야만 한다. 즉 인간이 보고 판단하는 것처럼 상황에 따라 동일한 물체라도 다른 카테고리로 판단하기 위해서는 그에 맞는 데이터정제가 이루어져야만 한다. 연구진의 시도결과 기존 라벨링 데이터를 수정하는 것보다 원천데이터에서 새로 라벨링을 하는 것이 오히려 작업량이 줄어드는 것을 확인하였다. 현 시점 개발이 진행 중인 인공지능 연구와 사업들은 대부분 위 문제점과 같이 하위 카테고리를 고려하지 않고 학습데이터를 수집, 정제하고 있기 때문에 후속 개발에서 활용되지 못하고 데이터 폐기물이 될 것으로 예상된다.

### 3.4 수집/정제/라벨링 관리자의 필요

수집/정제/라벨링 관리자의 개념은 특정분야의 전문지식을 보유하여 데이터가 학습에 유용할지 판단할 수 있는 인력으로 정의할 수 있다. 이 인력을 보유함으로써 얻는 이득은 개발기간의 단축과 정확도의 향상이다. 만일 관리자없이 정제가 이루어진 데이터로 학습된 모델의 성능이 낮다면 학습자료와 인공지능네트워크 둘 중 어디에서 문제가 발생했는지 원인을 찾아내기 어렵다. 이 경우 인공지능 연구자는 학습용 데이터 내에서 오류를 찾아내는 직무와 네트워크 개선 직무를 동시에 수행하게 되므로 연구개발 기간이 배 이상 늘어나게 된다. 하지만 데이터정제 관리자가 데이터 품질에 확신을 가지고 정제된 데이터를 제공한다면 인공지능 연구자는 순수하게 인공지능 네트워크 개발 업무에만 집중할 수 있을 것이며 이와 같은 분업화로 연구개발기간의 단축을 기대할 수 있을 것이다.

#### 4. 해결방안

3장에서 논의한 바와 같이 인공지능 연구개발에는 많은 인력자원이 소요된다. 특수분야 인공지능의 경우 그 분야의 고급교육을 받은 사람이 아니면 모델의 정확도를 확보할 수 없으며 이러한 수요의 인력은 항상 부족한 실태이다. 또한 당장의 개발과정에 맞춰 확장성을 고려하지 않은 데이터의 수집, 가공은 1회용품과 마찬가지로 한번 활용하고 버려질 것이다. 이와 같은 손실을 피하기 위해 개발초기부터 확장성을 고려하고 관리할 인력의 양성이 필요하다.

##### 4.1 인력의존적인 연구개발방식의 해결방안

대한민국은 국민의 상당수가 대학에 진학하며 컴퓨터 활용 능력은 세계최고 수준이다. 이에 인공지능연구개발에 필요한 일반적인 인적자원은 풍부하다고 할 수 있다. 하지만 턱없이 부족한 수의 인공지능 개발인력으로 인해 연구개발이 비효율적으로 이루어지고 있다. 어느정도 시간이 지나 인공지능이 대중화되는 시기에는 충분한 인력이 확보될 것이나, 4차산업혁명의 초기인 지금 이 시점에서 최고효율의 연구개발이 이루어져야 대한민국이 강국으로 도약할 수 있을 것이다.

결론적으로 인력의존적 문제는 인공지능 전문가에 의한 치밀한 학습방식의 선정, 데이터 수집, 정제, 라벨링을 통해 해결

할 수 있다. 추후 확장성까지 고려할 수 있도록 계획하는 것이 이상적이거나 전문가 혼자서 어려워지며 이에 따라 인공지능 전문가와 별도의 총괄관리자가 필요하다. 이와 같은 문제점을 개선하기 위해 Fig. 5와 같이 새로운 인공지능 개발과정의 인력 배분을 제안한다. 인력배분의 구조와 수치는 본 연구진의 경험적인 값이다. 총괄관리자는 본인의 전문분야와는 별개로 인공지능전문가의 계획을 이해하고 인력을 관리하는 업무를 수행하며 연구개발이 성공적으로 진행될 수 있도록 돕는 지원가 역할이다. 이러한 지원 인력이 부족하다면 최소 수집, 정제, 라벨링 단계의 관리자는 전문교육으로 선발하여 개발단계별로 관여하여 인력관리를 수행하고 인공지능 개발자는 모델의 성능을 높이는데 집중할 수 있도록 도와주어야 한다. 결국 산업혁명시대에 나타난 분업화와 다를 바가 없으며, 분업으로 생산성이 획기적으로 향상되었듯이 인공지능 연구개발도 동일한 성과를 보일 것으로 예상된다. 이를 위해 일반 연구자에게 적절한 교육을 통해 고급연구자로 성장시킬 수 있는 대규모의 교육프로그램과 지원이 필요하다.

##### 4.2 데이터 폐기물 발생우려의 해결방안

정부의 데이터 댐 구축사업을 통해 수집되는 1과제당 이미지의 수는 약 20만~50만장 수준이다. 이에 소요되는 비용은 18억원 내외이다(NIA, 2020) 과연 이 데이터를 다른 개발목적에

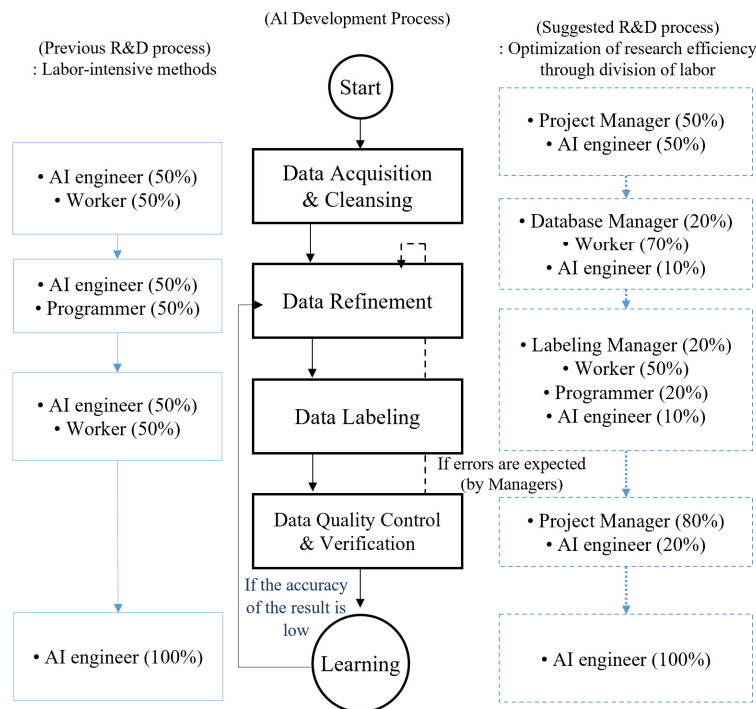


Fig. 5 The current R&D method for the human resource structure used in the AI development process and the improved manpower distribution method proposed by the researchers

적용할 수 있을 것인가는 재검토할 필요가 있다.

본 연구진의 사례연구에 따르면, 단순 분류 카테고리를 1가지 추가하는 데도 기존 정제량과 동일한 노력이 필요했다. 어떠한 인공지능 개발기법을 따르더라도 새로 개발하려는 모델에는 새로운 카테고리의 분류와 라벨링방식이 적용되며, 확장성을 고려하고 개발된 데이터셋이라 하더라도 기존의 절반 이상의 인력투입이 발생하였다. 따라서 연구개발을 시작하기 전 치밀한 카테고리 설정, 확장성을 고려한 데이터정리/라벨링 기술의 도입하거나 처음부터 개발범위를 넓게 시작하여 범용성을 가질 수 있도록 데이터를 수집, 정제하여야 하며 비용과 시간적 손실을 줄여야 할 것이다.

5. 사례연구와 작업지표의 제안

인공지능 연구개발이 일종의 경공업과 유사하다는 본 연구진의 주장을 뒷받침하기 위해 3가지의 사례연구를 수행하였다. 보통의 연구자들이 개발하는 방식과 유사하게 진행하였으며, 개발단계별로 발생하는 인력투입과 시간을 Table 4에서 확인할 수 있다. 여기서, 이미지 수량, 투입인력, 작업시간으로 다음과 같은 식 (1)을 제안한다.

$$\text{작업지표} = \frac{\text{데이터 총 수량(Amount)}}{\text{인력투입수준(Man)} \cdot \text{작업시간(hour)}} \quad (1)$$

이 식을 통해 데이터 대비 소모되는 자원의 수준을 정량적으로 계산할 수 있으며, 인공지능의 개발방식에 따라 1~10 사이의 값이 나타나는 것을 확인하였다. 데이터량 대비 투입된 인력과 시간이 많다면 지표는 낮은 값을 띄며, 추후 다른 연구개발과정에서 발생하는 데이터를 통계처리 한다면 데이터 수집과 정제, 라벨링 개발과정에 투입되는 자원이 적절하지 평가할 수 있을 것이다.

5.1 콘크리트 표면 균열인식 모델 개발

이미지분류기법(Object detection)을 사용하기 위해 2만장의 224x224pixel 사이즈의 콘크리트 균열 사진을 수집하였다.

콘크리트 균열은 실제 건축토목구조물에서 사진촬영을 통해 수집하였으며, 이미지 상에서 균열부분을 잘라내기하여 저장하였다. 데이터 수집과 잘라내기 작업에 약 1Man/hour로 200시간이 소요되었다. 그 외 특별한 라벨링작업이 없었기 때문에 많은 양의 학습데이터 대비 작업지표는 1.4로 나타났다. 학습결과 균열을 인식할 수 있었으나, 학습데이터 수집시 수평, 수직 균열에 위주로 모아져 대각균열은 인식하였으나 X 모양으로 교차되는 균열의 형태는 인식할 수 없었다.

5.2 건설폐기물 인식모델 개발

공사현장에서 발생하는 18종의 건설폐기물 중 5종인 콘크리트, 벽돌, 목재, 판재, 혼합폐기물을 구별하는 인공지능 모델을 개발하기 위해 이미지 분석(Instant Segmentation) 기법으로 데이터를 수집 정제하였다. 수집된 총 이미지는 866장으로 현장에서 직접촬영한 데이터와 온라인 상에 수집된 데이터를 512x512pixel 사이즈로 잘라내기 하였고, 장당 100kb 이내로 정제하였다. 또한 이미지 상에서 폐기물의 클래스를 구분하기 위해 물체 경계의 픽셀을 다각형(Polygon)방식으로 추출하였다. 데이터수집에는 3Man/hour로 48시간이 소요되었는데, 데이터수집 계획부터 현장에서 이미지를 촬영하고 이동시간, 데이터서버에 저장하는 시간을 모두 포함한 시간이다. 정제와 라벨링은 동시에 진행되어 4Man/hour로 180시간 소요되었다. Table 4와 같이 개발단계별로 작업지표가 상이하게 나타났으며, 이 연구개발의 경우 데이터수집에 많은 자원이 투입된 것을 알 수 있었다. 학습결과 정확도가 떨어지는 부분이 발생하는 현상이 나타나, 수집된 원천데이터를 바탕으로 재라벨링(세그멘테이션)에 4Man/hour로 60시간 정도 추가로 소요되었다. 재라벨링은 이전 작업에 비해 1/3의 자원이 투입되었으며 이렇게 재라벨링한 데이터로 학습결과 5종의 폐기물을 성공적으로 구분할 수 있었으나 판재와 목재를 혼동하는 사례가 발생하여 학습데이터량을 늘려야 할 것으로 판단되었다.

5.3 철근의 수량을 파악하는 모델 개발

건설현장에서 자동화된 자재관리를 위한 철근 분할인식모

Table 4 'Work index' on the basis of required manpower and hours depending on the AI R&D method

Case	Method	Amount of data	Step	Manpower/Hour	Working Time	Metrics per data	Avg.
A. Crack	Classification	20000	Acquisition/Labeling	1	200	1.4	1.4
B. Construction Wastes	Instant Segmentation	866	Acquisition	3	48	6.0	3.3
			Labeling	4	180	1.2	
			Re-labeling	4	60	3.6	
			Acquisition	3	48	5.0	4.1
Labeling	4	110	1.65				
C. Rebar		726					



델을 개발하였다. 512×512pixel 사이즈의 이미지 내에 50~250개의 철근다발이 촬영되었으며 날개의 철근을 분할인식하기 위해 철근의 단면을 폴리곤방식으로 추출하였다. 726장의 이미지를 수집하기 위해 3Man/hour로 48시간이 소요되었으며 약 12만개의 철근을 개별로 세그멘테이션 하기 위해 4Man/hour로 110시간 소요되었다. 학습 결과 철근의 단면을 분할인식하여 촬영된 이미지 내의 철근 개수가 추출되는 것을 확인했다. 다만, 한 장의 이미지 내에 300개 이상의 철근이 촬영되는 경우 512×512pixel 사이즈의 이미지로는 육안으로도 구분하기 힘든 철근이 다수 포함되어 있어 이미지 내의 해상도가 낮은 일부 영역에서 정확도가 현저히 낮아지는 것을 확인했다. 본 사례 연구의 경우 작업지표가 타 사례에 비해 높은 경향을 보이는데 이는 추가 라벨링 작업이 없었고 숙련된 연구팀에 의해 효율적으로 개발이 이루어졌기 때문으로 판단된다.

## 6. 결론

4차 산업혁명은 인공지능과 빅데이터, 블록체인, 로봇공학과 같은 여러가지 기술들이 동원된 지능화 사회를 말한다. 본 논문에서는 4차산업혁명 기술 중 인공지능에 중점을 두어 개발과정이 경공업의 과정과 유사함에 대해 연구하였다. 중공업에 비해 경공업은 인력이 많이 필요하며 생산단계/방식에 따라 효율이 결정된다. 인공지능 연구개발은 데이터 수집부터 정제, 학습, 품질평가에 이르기까지 모든 과정을 사람이 직접 수행하여야 하며 개발단계가 복잡한 것이 경공업과 매우 유사하다. 하지만 기존 사업체/기관에서는 인공지능 학습단계에만 중점을 두어 개발을 진행하였기 때문에 인력의 수요가 적다고 판단되어 왔으며, 이로 인해 국가연구개발과제개발 시에도 인력의 중요성이 낮게 평가되고 결국 제대로 된 인공지능 연구개발이 이루어지지 않고 있는 실정이다. 따라서 본 논문에서는 인공지능 학습단계별로 필요한 업무를 상세히 분석하였고 그 결과로 나타난 연구개발 인력양성의 시급성에 대해 알아보았다.

### 6.1 데이터 수집단계

기존에는 데이터 수집을 위해 단순히 이미지나 시계열데이터를 디지털자료로 복사를 하여 연구하였다 하지만 상용화목적, 특정전문분야 연구개발은 저작권, 개인신상정보 보호를 위해 기존 데이터를 마구잡이로 가져다 사용하는 것은 어려우며 수집 이후에도 개별 확인작업을 거쳐야 하는 과정이 필수적이다. 이 과정을 자동화해주는 툴이 존재하나 아직 완성단계라 볼 수 없기 때문에 전 작업에 사람의 개입이 필요하다.

### 6.2 데이터 정제/라벨링단계

정제기술의 고급화만으로도 인공지능 모델의 정확성과 연구기간의 단축을 기대할 수 있는 것은 사례연구를 통해 확인할 수 있었다. 이를 위해서는 개발분야의 전공지식을 보유한 정제관리자와 개발인력이 필요함이 나타났으며, 기존 클라우드 소싱 등으로 모집된 일반인들이 정제한 데이터가 아닌 개발분야 전문교육을 받은 인력이 정제한 데이터, 이를 관리하는 정제관리자를 고용하여 연구개발이 진행되어야 한다. 이를 통해 인공지능개발자는 본인의 업무에 집중할 수 있을 것이며 분업화의 효율성이 극대화 될 것이다.

### 6.3 품질평가 단계

연구결과에 따르면 품질평가는 각 단계의 관리자가 수행할 수 있을 것으로 보인다. 이러한 경우 연구목적에만 집중된 품질평가가 이루어질 수 있으며, 학습데이터의 재활용이 불가능한 경우가 발생한다. 현 시점에서 필수적이지는 않으나 인공지능연구의 황금기가 도달하였을 때 유사한 인공지능 연구개발을 위해 학습데이터의 재활용은 개발기간을 단축시키는 키가 될 수 있다. 이에 타분야개발까지 예상하여 데이터를 관리할 수 있는 전문인력의 필요성이 중요해질 것이다.

위 결과에 따라 충분한 인력자원을 교육하고 육성하여 분업화된 개발이 이루어진다면 인공지능 시대에 새로운 가치 창출이 이루어질 것으로 예상된다. 이로써 단계적으로 인공지능 산업은 인간의 일자리를 늘릴 것이며 어떠한 특이단계에 도달하기 전까지 인공지능으로 인하여 일자리가 감소되지는 않을 것으로 사료된다.

2차산업혁명 단계에서 나타난 분업과 기계장치의 발전은 폭발적인 생산력 향상을 가져왔다. 데이터라는 천연자원과 컴퓨터라는 기계장치로 인해 4차산업혁명은 2차산업혁명과 유사한 형태를 띠고 있다. 인공지능을 사용함으로써 늘어나는 생산력은 이미 여러 사례로 확인되었고, 이전에 그리하였듯이 인력을 효율적으로 배분하고 활용한다면 2차산업혁명 못 지 않는 인류의 발전을 기대할 수 있을 것이다.

## 감사의 글

본 연구는 한국과학재단이 주관하는 대학중점연구소지원사업(No. NRF-2018R1A6A1A07025819)과 신진연구지원사업(No. NRF-2020R1C1C1005406)의 지원을 받아 수행되었습니다.

## References

- Acuna, D., Ling, H., Kar, A., Fidler, S.** (2018) Efficient Interactive Annotation of Segmentation Datasets with Polygon-RNN++, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.859~868.
- Association, Telecommunications Technology** (2021a) AI Dataset Construction Guidebook (Ministry of Science and ICT).
- Association, Telecommunications Technology** (2021b) AI Learning Data Quality Management Guidelines v1.0.
- Baji, T.** (2018) Evolution of the GPU Device widely used in AI and Massive Parallel Processing, *IEEE 2nd Electron Devices Technology and Manufacturing Conference (EDTM)*, pp.7~9.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.** (2020) An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale, *Computer Vision and Pattern Recognition (cs.CV)*, arXiv preprint arXiv:2010.11929.
- Dutta, A., Gupta, A., Zissermann, A.** (2016) VGG Image Annotator (VIA), <http://www.robots.ox.ac.uk/~vgg/software/via>.
- Felzenszwalb, P., McAllester, D., Ramanan, D.** (2008) A Discriminatively Trained, Multiscale, Deformable Part Model, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1~8.
- Girshick, R.** (2015) Fast R-CNN, *Proc. IEEE International Conference on Computer Vision*, pp.1440~1448.
- Girshick, R., Donahue, J., Darrell, T., Malik, J.** (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.580~587.
- He, K., Gkioxari, G., Dollár, P., Girshick, R.** (2017) Mask R-CNN, *IEEE International Conference on Computer Vision*, pp.2961~2969.
- He, K., Zhang, X., Ren, S., Sun, J.** (2016) Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.770~778.
- Heo, S.J., Chunwei, Z., Yu, E.J.** (2018) Response Simulation, Data Cleansing and Restoration of Dynamic and Static Measurements Based on Deep Learning Algorithms, *Int. J. Concr. Struct. Mater.*, 12, pp.1~13.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.** (2017) Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.4700~4708.
- Kim, B.S., Lee, J.H., Kang, J.W., Kim, E.S., Kim, H.J.** (2021) HOTR: End-to-End Human-Object Interaction Detection with Transformers, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.74~83.
- Kim, J.M., Hyeon, S.G., Chae, J.H., Do, M.S.** (2019) Road Crack Detection based on Object Detection Algorithm using Unmanned Aerial Vehicle Image, *J. Korea Inst. Intell. Trans. Syst.*, 18, pp.155~163.
- King, T.M., Arbon, J., Santiago, D., Adamo, D., Chin, W., Chin, W., Shanmugam, R.** (2019) AI for Testing Today and Tomorrow: Industry Perspectives, *IEEE International Conference On Artificial Intelligence Testing (AITest)*, pp.81~88.
- Krizhevsky, A., Sutskever, I., Hinton, G.E.** (2012) Imagenet Classification with Deep Convolutional Neural Networks, *Adv. Neural Inf. Proc. Syst.*, 25, pp.1097~1105.
- Lee, B.Y., Yi, S.T., Kim, J.K.** (2007) Surface Crack Evaluation Method in Concrete Structures, *J. Korean Soc. Nondestruct. Test.*, 27, pp.173~182.
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L.** (2014) Microsoft COCO: Common Objects in Context, *European Conference on Computer Vision*, Springer, pp.740~755.
- Lowe, D.G.** (1999) Object Recognition from Local Scale-Invariant Features, *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp.1150~1157.
- Na, S.G., Heo, S.J., Han, S.H.** (2021) Construction Waste Reduction through Application of Different Structural Systems for the Slab in a Commercial Building: A South Korean Case, *Appl. Sci.*, 11(3), p.5870.
- NIA (National Information Society Agency)** (2020) ‘NIA2020-032’, <https://www.nia.or.kr>.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A.** (2016) You Only Look Once: Unified, Real-Time Object Detection, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.779~788.
- Ren, S., He, K., Girshick, R., Sun, J.** (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *Adv. Neural Inf. Proc. Syst.*, 28, pp.91~99.
- Revaud, J., Heo, M., Rezende, R.S., You, C., Jeong, S.G.** (2019) Did It Change? Learning to Detect Point-of-Interest Changes for Proactive Map Updates, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.4086~4095.
- Roh, B.S., Shin, W.H., Kim, I.D., Kim, S.W.** (2021) Spatially Consistent Representation Learning, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.1144~1153.
- Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.** (2008) LabelMe: A Database and Web-Based Tool for Image Annotation, *Int. J. Comput. Vis.*, 77(1-3), pp.157~173.

**Simonyan, K., Zisserman, A.** (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition, *Computer Vision and Pattern Recognition*, arXiv:1409.1556.

**Tan, M., Le, Q.** (2019) Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks, *International Conference on Machine Learning*, pp.6105~6114. PMLR.

**Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.** (2017) Attention is All You Need, *31st Conference on Neural Information Processing Systems(NIPS 2017)*, pp.5998~6008.

---

## 요 지

본 연구는 인공지능 연구개발과정에 많은 인적자원이 필요함을 인지하고 현 개발방식 고려할 사항에 대해 논의한다. 결론적으로 인공지능 개발의 효율성 향상을 위해서는 소수의 관리자와 많은 일반작업자들의 분업화가 이루어져야 가능하며, 이는 마치 일종의 경공업의 형태와 유사하다고 생각된다. 따라서 본 연구진은 컴퓨터라는 기계장치로 데이터라는 디지털 자원을 다루어 생산의 효율성을 높이는 인공지능 개발과정을 4차산업시대의 디지털 경공업이라고 명명한다. 이전 산업혁명시대에서 경험한 것과 마찬가지로 인적자원을 효율적으로 배분화하고 활용한다면 디지털 경공업은 2차산업혁명 못 지 않는 발전을 기대할 수 있을 것이며, 이를 위한 인력양성이 시급하다고 판단된다.

**핵심용어** : 디지털 경공업, 4차산업혁명, 인공지능, 인력양성, 작업지표, 데이터폐기물