

자동-레이블링 기반 영상 학습데이터 제작 시스템

An Auto-Labeling based Smart Image Annotation System

이용, 장래영, 박민우, 이건우, 최명석
한국과학기술정보연구원 연구데이터공유센터

Ryong Lee(ryonglee@kisti.re.kr), Rae-young Jang(raezero@kisti.re.kr),
Min-woo Park(pminwoo@kisti.re.kr), Gunwoo Lee(gwlee@kisti.re.kr),
Myung-Seok Choi(mschoi@kisti.re.kr)

요약

최근 딥러닝 기술의 급속한 발전과 함께 학습데이터가 크게 주목을 받고 있다. 일반적으로 딥러닝 방식에서는 모델을 훈련시키기 위해 충분한 학습데이터가 준비되어 있어야 한다. 하지만, 딥러닝 모델 설계 작업과 달리 데이터셋을 제작하는 데 상당한 시간과 노력이 필요하다. 영상 데이터를 주로 다루는 시각지능 분야에서도 학습데이터 제작자들은 전문적인 학습데이터 제작 도구를 사용해 이미지 단위로 레이블링을 수작업으로 하고 있어 여전히 많은 시간과 노력이 필요한 상황이다. 따라서, 다양한 분야에서 필요한 충분한 영상 학습데이터셋을 확보하기 위해 기존의 수작업 방식을 대체할 수 있는 레이블링 기술이 필요하다. 본 논문에서는, 영상 학습데이터셋 동향을 소개하고, 학습데이터 제작 환경에 대해 분석한다 특히, 수작업으로 이루어지는 반복적이고 수고스러운 레이블링 과정을 자동화하여, '확인과 수정'의 단계를 비약적으로 단축시킬 수 있는 '스마트 영상 학습데이터 제작 시스템'을 제안한다. 그리고, 실험을 통해 영상 학습데이터 제작 과정에서 이미지에 박스형 및 폴리곤형 객체영역을 지정하여 레이블링하는 데 소요되는 시간을 크게 줄이기 위한 자동레이블링 방식의 효과를 검증한다. 마지막으로, 제안하는 시스템의 실험에서 추가적으로 검증되어야 하는 부분과 함께 이를 개선하기 위한 향후 연구 계획에 대해 논의한다.

■ 중심어 : | 영상 학습데이터셋 | 이미지 레이블링 도구 | 자동 레이블링 |

Abstract

The drastic advance of recent deep learning technologies is heavily dependent on training datasets which are essential to train models by themselves with less human efforts. In comparison with the work to design deep learning models, preparing datasets is a long haul; at the moment, in the domain of vision intelligent, datasets are still being made by handwork requiring a lot of time and efforts, where workers need to directly make labels on each image usually with GUI-based labeling tools. In this paper, we overview the current status of vision datasets focusing on what datasets are being shared and how they are prepared with various labeling tools. Particularly, in order to relieve the repetitive and tiring labeling work, we present an interactive smart image annotating system with which the annotation work can be transformed from the direct human-only manual labeling to a correction-after-checking by means of a support of automatic labeling. In an experiment, we show that automatic labeling can greatly improve the productivity of datasets especially reducing time and efforts to specify regions of objects found in images. Finally, we discuss critical issues that we faced in the experiment to our annotation system and describe future work to raise the productivity of image datasets creation for accelerating AI technology.

■ keyword : | Image Training Dataset | Image Labeling Tool | Automatic Labeling |

* 본 논문은 ICC 2020 국제학술대회 우수논문입니다.

* 본 연구는 한국과학기술정보연구원(KISTI) 'Data/AI 기반 문제해결 체계 구축(K-21-L01-C03)'사업 지원으로 수행되었습니다.

접수일자 : 2021년 04월 09일

심사완료일 : 2021년 05월 25일

수정일자 : 2021년 05월 25일

교신저자 : 최명석, e-mail : mschoi@kisti.re.kr

I. 서론

인공지능 기술의 실용적 수준의 발전으로 다양한 분야에서 이를 활용하고자 하는 요구가 크게 급증하고 있다. 이러한 요구에 발맞추어, 실제적으로 활용 가능한 수준의 인공지능 기술로써 딥러닝으로 대표되는 데이터 기반 모델 훈련 방식이 개발에 사용되고 있다. 이러한 딥러닝 기반의 인공지능을 개발하기 위한 소프트웨어 및 하드웨어 기술은 글로벌 IT 기업들의 리딩으로 크게 발전하고 있는 반면, 딥러닝 모델을 실제적으로 훈련시키기 위해 반드시 필요한 데이터, 소위 학습데이터에 대한 중요성은 모델 개발 기술에 비해 덜 중요하게 여겨지고 있는 상황이다. Cognilytica 조사기관[1]에 따르면 일반적인 AI 프로젝트에서 학습데이터의 준비과정이 전체 프로젝트의 80% 시간을 소모하고 있으며, 향후 이러한 데이터의 수요와 더불어 그 시장이 2023년까지 \$1.2B에 이를 것으로 예상하고 있다. 특히, 학습데이터 제작을 위한 레이블링 시장은 급증하는 수요에 대응하기 위해 크게 확대되며 2023년까지 \$1B 까지 확대될 것으로 예상하고 있다.

학습데이터는 딥러닝 기반의 인공지능 개발에서 전문가에 의해 설계된 모델을 훈련시키는 교재와 같은 역할을 담당한다. 여기서, 학습데이터는 모델 설계 및 인프라 구축과 달리, 인공지능이 해결해야 하는 문제별로 필요에 따라 제작해야 한다. 하지만, 일반적으로 사람에게 의해 직접 만들고 있어 단기간에 대량으로 확보하기 어렵다는 문제를 안고 있다. 다시 말해, 인공지능 연구개발자들은 소규모 단위로 학습데이터를 필요에 따라 만들고 있는 실정으로, 이러한 상황은 연구생산성의 저하를 야기시킨다. 이러한 문제점을 극복하기 위해, 기업들은 인공지능 기술 개발을 위한 데이터 확보에 지속적으로 많은 비용을 지출하고 있으며, 국가적으로는 다양한 사회현안 해결 등에 필요한 학습데이터를 주도적으로 생산하여 공유하고 있다. 하지만, 모든 분야에 필요한 데이터를 직접적으로 제작하여 제공하는 것은 현실적으로 불가능하기 때문에, 지속적으로 학습데이터를 효과적으로 제작하고 공급할 수 있는 솔루션 개발이 필요하다. 따라서, 학습데이터 제작 솔루션은 인공지능 기술 개발의 핵심적인 기술의 한 부분으로서 연구자 및 기업

은 물론 국가적으로도 중요한 미래 국가경쟁력 기반이 될 수 있다.

본 논문에서는 시각지능 개발에 필요한 학습데이터 제작을 효과적으로 지원하기 위한 시스템을 소개한다. 제안하는 시스템에서는 자동 레이블링으로 도출되는 결과에 대한 확인과 수정에만 작업자가 관여하여, 효과적이고 빠른 학습데이터 양산이 가능하게 된다. 자동 레이블링은 학습데이터 제작 시간을 크게 단축 시킬 수 있는 기술로서, 다양한 레이블링 전문 기업에서 경쟁적으로 기술 개발 및 적용을 위한 노력을 진행하고 있다. 본 연구에서는 기존의 학습데이터 제작 서비스들이 일반적으로 한정된 사전학습 모델에 의존하는 자동레이블을 적용하는 것과 달리 백-엔드 온라인 인공지능 서비스와 연동하여 보다 다양한 사전학습 모델을 적용할 수 있는 기술을 제안한다.

본 논문의 구성은 다음과 같다. 우선, 2장에서 영상 학습데이터 확보를 위한 개인 연구자부터 국가 레벨의 다양한 노력을 살펴본다. 그리고, 이미지 레이블링을 위한 온/오프라인 레이블링 작업 환경들의 비교·분석을 통해 학습데이터 제작 생산성 제고를 위한 전략들을 검토한다. 3장에서는 본 논문에서 제안하는 백-엔드 온라인 인공지능 서비스 기반 영상 학습데이터 제작 지원 시스템 DALDA에 대해 구체적으로 소개한다. 4장에서는 제안 시스템을 통해 학습데이터 제작 시간 단축 효과를 실험적으로 검증한다. 그리고, 5장에서 실험 과정에서 경험한 제약과 한계와 함께, 이를 해결하기 위한 향후 연구 계획에 대해서 설명하며, 6장에서 본 연구의 요약으로 끝을 맺는다.

II. 관련 연구

본 장에서는 영상 인식 분야에서 많이 활용되는 학습데이터셋과 제작환경에 대해 비교 설명한다. 특히, 학습데이터 제작에는 많은 노력과 비용이 수반되기 때문에 오래된 이전의 많은 데이터셋이 아직까지 활용되고 있기도 하다. 학습데이터 제작환경은 인공지능 기술확산과 함께 소규모 제작 프로그램에서 온라인 제작 서비스로 비즈니스화하는 형태로 진화하고 있다.

1. 영상 학습데이터셋 동향 소개

다양한 영상인식 기술의 발전과 함께 실제 인식대상 에 대한 특징추출 및 인식평가 등의 목적으로 학습데이터의 제작과 공유는 연구커뮤니티를 중심으로 진행되어 왔으나, 더 많은 분야에서 인공지능 기술 활용을 위해 학습데이터 제작의 필요성이 강조되고 있다. 일반적으로, 인공지능 개발을 위한 학습데이터의 제작 형태는 다음과 같은 몇 가지 형태로 나눌 수 있다.

1) 소규모 연구자 중심[2][3][4] : 시각지능 연구를 수행하는 연구자가 필요에 따라 적절한 영상 데이터들을 수작업으로 수집/선별하여 제작하는 경우이다. 개인 또는 소그룹으로 진행하니 큰 규모로 발전시키기 어렵고, 이미지 분류와 같은 작업에서 주관적인 관점의 선택이라는 한계를 가질 수 있다.

2) 연구 커뮤니티 중심[5-7] : 다수 연구자들에 의해 온라인으로 공동으로 데이터 수집, 레이블링하여 학습데이터를 제작하며 지속적인 검토작업을 통해 데이터의 품질을 개선한다. 또한, 수집 데이터 규모나 품질 면에서 소규모 연구자 중심으로 개선된 결과를 얻을 수 있지만, 공통된 관심을 갖는 다수 연구자들의 커뮤니티를 구성하고 지속적으로 유지해야 하는 어려움이 있다.

3) 전문 학습데이터 제작 기업형[8-10] : 최근에 증가하는 학습데이터 수요에 맞추어 원시 영상데이터 수집부터 학습데이터를 제작하는 기업들이 등장하고 있다. 특히, 학습데이터를 전문적으로 제작하는 소프트웨어나 다수의 작업자를 강점으로 단기간에 집중력 있게 대량의 학습데이터를 만들어 내고 있다. 하지만, 비전문가가 학습데이터 제작에 숙련되는 시간이 소요되며 연구커뮤니티와 다르게 다수의 작업자간의 공통된 의견으로 일관성있게 고품질의 학습데이터를 제작하는 게 어렵다는 한계를 가지고 있다.

4) 온라인 대규모 참여형 : 구글의 Open Images Dataset[11]처럼 대규모 영상 이미지를 대상으로 학습데이터를 수집하고 검색, 공유 서비스를 제공하는 온라인 서비스가 등장하고 있다[12][13]. Kaggle[14]과 같

이 데이터 분석 및 인공지능 모델 개발 등의 경진대회를 중심으로 다수의 학습데이터가 자연스럽게 공유되는 서비스를 통해서도 지속적인 데이터 수집을 할 수 있다. 이러한 글로벌 데이터 공유 서비스를 중심으로 향후 필요한 클래스의 이미지 또는 객체를 포함한 이미지를 손쉽게 검색하고 활용할 수 있는 서비스가 증가할 것으로 예상된다.

5) 국가 주도형 : 인공지능 기술은 점차 국가 기술력을 가늠하는 핵심이 되고 있으며, 그 가운데 있는 학습데이터 구축은 이를 지지하기 위한 기반이 되고 있다. 연구자나 개발자 등이 원활하게 데이터를 확보하기 어렵게 되면 그만큼 인공지능 기술개발이 더디어지기 때문에 선진 각국은 국가 주도로 데이터를 제작, 공개하는 노력을 기울이고 있다. AI Hub[15] 사업을 주축으로 국내에서도 다양한 분야의 영상 학습데이터 니즈를 충족시키기위해 노력하고 있다. 하지만, 지속적으로 늘어나는 다양성과 양적 수요에 신속히 그리고 충분히 대응하기 어려운 한계를 갖는다. 따라서, 정부 주도의 분야별 대규모 학습데이터 제작 노력과 더불어 데이터를 손쉽게 빠르게 제작할 수 있는 기술개발이 시급하다.

6) 데이터 증식 기술을 통한 유사 학습데이터 양산 : GAN[16][17]으로 대표되는 이미지 생성기술을 통해 흥미로운 영상을 자동으로 만들어 내어 예술 분야 등에서 주목을 모으고 있다. 한편, 실제로 데이터를 수집하기 어려운 환경의 영상을 생성할 수 있는 수준으로 발전하고 있으며, 이러한 기술을 기반으로 부족한 학습데이터를 만들어 내어 딥러닝 모델을 훈련시키는 데 사용할 수 있다.

[표 1]에서는 시각지능 분야에서 많이 사용되는 주요 영상 학습데이터셋을 비교하고 있다. 기본적으로 이미지 분류, 객체인식, 캡션 등 각기 다른 목적으로 제작되어 공유되어 왔다. MNIST, CIFAR 데이터셋과 같이 초기의 딥러닝 기술 개발 시에 활용되었던 데이터셋은 그 규모가 작고, 클래스도 적고 다양성도 적어 주로 연구 목적으로만 이용되어 왔다. 현재는 자율주행을 위한 데이터셋인 KITTI나 Google Images Dataset과 같이 그 다양성과 에서도 크게 확대되고 있는 것을 알 수 있다.

표 1. 주요 영상 학습데이터셋의 비교

	타겟 분야	목적	규모	레이블 표현	레이블 구조	제작 주체
MNIST[2]	숫자인식	분류	10 클래스, 60,000장 이미지 (클래스별 6,000장), 28x28	자체 Binary 포맷	Image - Label	LeCun
CIFAR-10[3]	영상인식	분류	10 클래스, 60,000장 이미지 (클래스별 6,000장), 32x32	자체 Binary 포맷	Image - Label	Krizhevsky, et. al
CIFAR-100[3]	영상인식	분류	100 클래스, 60,000장 이미지 (클래스별 600장), 32x32	자체 Binary 포맷	Image - Label	Krizhevsky, et. al
PASCAL VOC[5]	영상인식	분류/ 객체탐지	20 classes, 9,963 images	XML	Image-Label Image-Region-Label	Mark Everingham, et. al
MS COCO[6]	영상인식	분류/ 객체탐지/ 캡션	100 클래스, 2.5 million instances/ 328k이미지	JSON	Image-Label Image-Region-Label	COCO Consortium
ImageNet[7]	영상인식	분류/ 객체탐지	21841 synsets, 14,197,122 images	CSV, WordNet 기반	Image-Label Image-Region-Label	ImageNet Project
KITTI[8]	자율주행	객체탐지	Images, GPS/IMU, Velodyne / 180GB	Complex	3D Lidar, Image, Sensor Data	KIT
Caltech 256[18]	영상인식	분류	256 classes, 30,000 images	Folder	Image - Label	Griffin, et. al
Cityscapes[9]	스마트 시티	객체탐지	5,000 frames (HQ), 20,000 frames (WA)	JSON	Image - Polygon	Cityscape
IMDB[10]	얼굴 인식	분류	얼굴 이미지: 460,723, 클래스: 20,284	매트랩 저장 파일 (.mat)	Image - Label	IMDB, Rasmus, et. al
Kinetics[12]	동작 인식	분류	클래스: 700, 650,000 video clips	CSV	Video Time-stamp - Label	DeepMind
MPII Human Pose[4]	동작인식	분류	410 human activities/ 25k images, 40 people	매트랩 저장 파일 (.mat)	Image - Label	Andriluka, et. al
Google Open Images Dataset + V6 [11]	영상 인식	분류/ 객체탐지/ 캡션	19,957 classes, 59,919,574 images	CSV	Image - Region - Caption	Google
20BN[13]	동작인식	분류/캡션	174 classes, 220,847 videos	JSON	Video - Label	TwentyBN (Commerical)
Conceptual Captions[15]	영상인식	캡션	2,007,528 이미지와 설명문	CSV	Image - Text	Google

인공지능 개발에서 데이터의 중요성에 대한 인식과 더불어 많은 투자, 군중 참여 등으로 데이터 제작 활동이 크게 활성화되고 있다. 또한, 영상 분야의 주된 콘텐츠인 이미지와 비디오를 중심으로 레이블링 정보가 연구 고유의 자체 포맷에서 CSV, XML, JSON 등 공유 활용 촉진을 위한 형태로 바뀌어 오고 있으며, OpenAPI 형태로도 데이터를 취득할 수 있게 되고 있다. 그럼에도 불구하고, 다양한 분야에서 만들어진 원시

영상데이터를 학습데이터로 바꾸기 위해서는 많은 레이블링 작업자들이 각각의 데이터를 정리하고, 개별 데이터의 내용에 따라 레이블링 작업에 부담이 있다.

2. 영상 학습데이터 제작 환경 분석

영상 학습데이터 제작을 위해서는 아직도 많은 부분이 있어 사람의 수작업이 필수적으로 수반된다. 특히, 딥러닝 모델이 학습데이터를 통해 훈련된다는 속성 때

문에 적절한 학습데이터의 확보가 중요하다. 실제적으로, 작업자가 직접 이미지를 보고 적절한 레이블링 작업을 수행하고, 딥러닝 모델이 이해할 수 있는 형태로 만들어야 한다. 영상 학습데이터 제작에 있어 레이블링은 크게 몇 가지 형태로 나눌 수 있다.

1) 이미지 단위 분류 레이블링 : 이미지 단위로 객체 또는 다양한 상황을 사전에 정해진 카테고리 범위에서 분류하기 위한 데이터셋 제작을 목적으로 이미지 자체에 클래스명을 매핑하는 수준이다[2]. 학습데이터 제작은 이미지를 보여주고 클래스명에서 선택하게 하는 형태로 진행되며 학습데이터 제작 도구도 가장 단순하게 구성될 수 있으며, Google의 Image Open Images Dataset[11] 사이트에서는 온라인으로 군중참여형으로 태그를 부여할 수 있다.

2) 이미지 내 객체 레이블링 : 이미지 내의 다양한 객체를 인식하기 위한 데이터셋으로 레이블링 작업자에게 이미지를 보여주고 객체영역과 해당 객체의 클래스를 선택하는 작업을 필요로 한다[20][21]. 영역지정은 일반적으로 바운딩 박스 또는 바운딩 폴리곤 형태로 지정하게 하며, 작업환경 소프트웨어가 이러한 작업을 GUI를 통해 지원해야 한다. 바운딩 박스는 객체 출현 영역을 사각형의 영역을 비교적 단순히 지정할 수 있는 반면, 폴리곤 영역은 다수의 점으로 구성된 폐쇄 영역을 지정해야 해서 작업에 상당한 시간이 소요되며 작업자별로 폴리곤을 표현하는 방식이 크게 다를 수 있는 문제점도 있다.

3) 이미지에 대한 설명 레이블링 : 이미지에 대한 사람 수준의 이해를 텍스트로 표현한 것으로 이미지 단위로 설명하는 텍스트의 쌍으로 레이블링을 포함한다. Google의 Conceptual Caption 데이터셋은[15] 2백만장의 이미지와 이에 대한 간단한 설명문을 제공하고 있다. 주로 이미지의 출현 객체와 상황을 묘사하기 위한 텍스트를 표현한 것으로 작업환경과 작업방식은 비교적 단순하지만 작업자에게 이미지를 텍스트로 묘사하게 해야 해서 개별 작업자들의 주관적 관점의 한계를 가지게 된다.

상기와 같이 이미지를 인식하고 활용하기 위한 목적별로 적절한 인공지능 모델이 개발되어야 하며 주로 모델이 원하는 출력형태와 동일한 구조의 학습데이터를 제작해야 한다.

학습데이터를 제작하는 제작환경 및 작업형태를 살펴보면, 상기 레이블링 방식 중 1)과 3)은 비교적 단순히 이미지를 보고 직접 선택하거나 설명문으로 표현하는데 반해, 2)의 객체 레이블링은 이미지 내의 객체 영역과 클래스 지정을 직접해 주어야 해서 전문적인 저작 도구 활용이 불가피하다.

객체단위의 이미지 레이블링을 위한 이미지 어노테이션 또는 이미지 레이블링 툴이라고 불리는 저작도구들이 등장하고 있으며 학습데이터 수요 증가와 함께 주목받고 있다. 현재 공개된 이미지 학습데이터 저작도구들의 비교는 [표 2]와 같다. 이들 제작환경들은 영상 학습데이터의 수요와 함께 개발되어 활발하게 활용되고 있는 것으로, 일부는 유상으로 서비스를 제공하고 있다. 크게 오프라인과 온라인 방식으로 나뉘며, 제작 소프트웨어를 설치해야 하는 오프라인 방식과 비교해, 사이트로 접속해서 하는 온라인 방식은 웹 브라우저를 통해 작업할 수 있는 장점이 있다. 하지만, 대량의 이미지나 대용량 비디오 상에서 작업을 수행해야 하며 오프라인 방식이 상대적으로 편리하다. 대상으로 하는 데이터는 주로 이미지와 동영상이며, 최근 자율주행을 위한 인공지능 기술 개발에 발 맞추어 3D Point Cloud 데이터를 지원하는 경우도 있다. 출력형식은 주로 VOC, COCO와 같은 형식을 지원하여 최종 데이터셋을 바로 모델 훈련에 사용할 수 있도록 하고 있다. 또한, 얼굴 등 개인정보를 파악할 수 있는 객체를 지우기 위한 모자이크 기능이나 다수의 작업자들을 통한 협업을 지원하기 위한 작업 파이프라인 설계를 지원하는 서비스도 있다. 특히, 자동 레이블링 기능은 작업자의 레이블링 전에 사전학습된 인식 모델을 통해 객체를 인식하고 작업자에게 이후 작업을 수행하게 하도록 하고 있다. 본문에서 제안하는 자동 레이블링 기능도 이러한 기능의 일종이지만, 기존 작업환경 소프트웨어/서비스와 달리 사전 학습모델을 레이블링 작업환경과 일체화시키는 것이 아닌 별도의 백-엔드 온라인 인공지능 서비스를 통해 제공한다는 점에서 차별화된다.

표 2. 영상 학습데이터셋 제작 도구 비교

	작업환경	처리 대상	지원 출력형식	자동 레이블링	상용 여부	특징
LabelMe[20]	오프라인	이미지, 비디오	VOC, COCO	미지원	무상	
MS VoTT[21]	온/오프라인	이미지, 비디오	Azure., CNTK VOC, COCO, CSV, 자체	지원	무상	
Make-Sense[22]	온라인	이미지	YOLO, VOC, COCO	지원	무상	
SuperAnnotate[23]	온라인	이미지	JSON	지원	유상	
VIA(VGG)[24]	온라인	이미지, 비디오	JSON, CSV	미지원	무상	
Supervise.ly[25]	온라인	이미지, 비디오, Point Cloud	JSON	지원	유상	
Labelbox[26]	온라인	이미지	JSON, CSV	지원	유상	
Hasty.ai[27]	온라인	이미지	VOC, COCO, Json	지원	유상	
CVAT[28]	온라인 /다중 유저	이미지	VOC, COCO 등	지원	유상	작업관리
Darwin[29]	온라인	이미지, 비디오	VOC, COCO, JSON	지원	유상	DICOM 지원
Heartex[30]	온라인 /다중 유저	이미지, 비디오, 텍스트, 오디오	VOC, COCO, JSON, CSV	지원	무상	
Scalabel[31]	온라인 /다중 유저	이미지, 비디오, Point Cloud	자체 형식	지원	무상	
Segments.ai[32]	온라인	이미지	COCO	지원	유상	
LabelIMG[33]	오프라인	이미지	VOC, YOLO	없음	무상	
ImageTagger[34]	온라인 /다중 유저	이미지	사용자 설정	없음	무상	
DarkLabel[35]	오프라인	이미지, 비디오	VOC, YOLO	지원: 연속 프레임에서 객체 추적 기능	무상	모자이크 지원
EVA[36]	오프라인	비디오	VOC, YOLO	지원: 연속 프레임에서 객체 추적 기능	무상	모자이크 지원
LOST[37]	온라인/다중 유저	이미지	CSV	지원	무상	작업 파이프라인 설계 지원
Ybat[38]	온라인	이미지	VOC, YOLO, COCO	없음	무상	
MuViLab[39]	오프라인	비디오	JSON	없음	무상	이벤트 단위 처리
Turkey[40]	온라인	이미지	CSV	없음	무상	Amazon Mechanical Turk 기반
Point Cloud Annotation Tool[41]	오프라인	Point Cloud (3D Points)	Applo 3D	없음	무상	포인트 클라우드
DeepLabel[42]	오프라인	이미지	KITTI, VOC, COCO, YOULO	지원	무상	

이를 통해, 다양한 사전학습 모델로 확장 가능하며, 작업자는 사전학습 모델을 구동시키기 위한 별도의 컴퓨팅 준비가 필요 없으며 다수의 작업자가 백-엔드 서비스를 공유해서 사용할 수 있어 **전체적으로 효율적인 자동 레이블링 작업 환경을 구축할 수 있다.**

III. 인공지능 기반의 스마트 영상 학습 데이터 제작 시스템 (DALDA)

영상 학습데이터 제작에서 가장 시간과 노력이 많이 요구되는 객체 단위 레이블링 작업을 지원하기 위해 본



그림 1. 스마트 이미지 레이블링 시스템 DALDA의 자동 레이블링 작업 흐름도

연구에서 딥러닝 기반의 자동 레이블링 기능을 갖는 스마트 학습데이터 제작 시스템을 개발하였다. 특히, 기존의 작업자들이 가장 큰 부담을 느끼는 부분인 이미지에서 객체 영역을 지정하는 작업의 부담을 줄이고, 가능한 한 다양한 분야에서 활용할 수 있도록 [그림 1]과 같이 사전 레이블링 작업에 활용될 수 있는 다양한 딥러닝 모델을 적용할 수 있는 시스템을 목표로 하였다. 특히, 이미지 레이블링 시스템을 구성하기 위해 오픈소스인 MS VoTT[21]를 기반으로 기능 확장을 통해 진행하여, 기본 기능인 이미지/비디오 로딩 및 GUI를 통한 바운딩 박스 및 폴리곤에 의한 객체 영역 지정 등을 바로 이용할 수 있었다.

본 연구에서는 자동 레이블링 기능을 중심으로 개발을 추진하였으며, 구체적으로 스마트 도시 분야에서 많이 생산되는 도시 영상 데이터를 활용해 학습데이터 제작을 지원하기 위한 필요한 기능을 부가적으로 개발하였다. 본 장에서는 영상 학습데이터 제작에서 필요한 수작업을 줄이기 위한 자동 레이블링의 실제 구현 방식 및 백-엔드 온라인 객체 인식 서비스 기능에 대해 설명한다.

1. 인공지능 기반 자동 레이블링

대부분의 이미지 레이블링 작업 S/W 또는 서비스는 공통적으로 작업자에게 이미지를 직접 보여주고, 판단하게 하며 이미지 또는 객체 단위로 레이블링을 수행하게 만든다. 인공지능 시대에 사람이 이렇게 수동적으로 레이블링 작업을 수행해야 하는 것이 모순적인 것처럼 보이지만 학습데이터를 기반으로 모델을 훈련시키는 딥러닝의 기본방식 때문에 피할 수 없는 과정이다. 특히, 다수의 객체를 더 정확하게 인식하기 위해서는 보통 객체별로 수백~수천장 이상의 사진을 필요로 하기 때문에 원시 데이터 확보 뿐만 아니라 이에 레이블링을 하고, 학습데이터로 정리해내는 작업 또한 중요한 부분이다. 사람의 직접적인 노동을 줄이기 위해 작업물은 더 편리하게 만들어질 필요가 있으며 반복적인 작업을 줄이는 방향으로 개선시킬 필요가 있다. 특히, 객체 출현 영역을 지정하는 작업에 시간이 많이 소요되며, 폴리곤 영역으로 정교하게 객체를 지정하려면 폴리곤의 영역을 점 단위로 지정해가면서 작업해야 해서 더 많은 시간을 필요로 한다. 이러한 수고를 줄이기 위해 객체를 우선 사전에 학습된 모델을 통해 찾게 만들고 그 결과를 사람에 의해 검증하고 수정하게 만드는 방법으로 개선할 수 있다.

자동 레이블링이라는 이 방식은 [표 2]에서와 같이 현재 다수의 이미지 레이블링 서비스에서 경쟁적으로 도입 중에 있으며, 본 제안 시스템에도 이 방식을 보다 개선된 방식으로 적용하고 있다. 특히, 일반적인 자동 레이블링 방식은 사전에 특정 사전학습 모델을 내장시켜 이를 통해 작업자에게 이미지에 대한 사전 레이블링 결과를 보여주고 이를 검수, 수정하는 형태로 작업을 하게 한다. 하지만, 본 제안 시스템에서는 보다 확장성 있는 자동 레이블링을 위해 다양한 사전학습모델을 적용할 수 있도록 하고 있다. 이를 위해 학습모델을 시스템에 내장하는 방식이 아닌 백-엔드 온라인 객체 인식 방식으로 구성하였다.

· **내장 사전인식모델 방식** : [표 2]에서 자동레이블링을 지원하는 제작환경들이 모두 이 방식에 속한다. 인식시스템에 고정된 모델을 사용하며, 작업자 환경에 전용 소프트웨어를 설치하는 방식의 경우 인식모델 구동을 위한 사전 소프트웨어 설치 및 구동을 위한 별도의 컴퓨팅 파워를 필요로 하는 부담이 있다. 또한, 온라인 방식이라도 고정된 사전인식 모델만을 사용한다는 한계를 갖는다.

· **백-엔드 온라인 인식 방식** : 작업자가 레이블링 작업을 하는 소프트웨어 자체는 온/오프라인 어느 방식으로든 가능하나 자동 레이블링의 핵심이 되는 사전인식 모델 기반 객체검출부는 온라인으로 서버에서 처리하게 만들어 작업자 환경에 부담을 줄일 수 있다. 특히, 서버의 인식기능을 다수의 작업자가 공용으로 사용하여 비용부담을 줄이고 사용성을 높일 수 있다. 또한, 인식모델을 필요 시에 손쉽게 업데이트하고 교체할 수 있어 학습데이터가 누적됨에 따라 보다 사전인식이 성능이 개선되고 작업자는 점차 수정보다 검증작업으로 작업부담을 크게 줄일 수 있다.

본 연구에서는 이미지 레이블링 시스템에서 자동 레이블링을 수행하는 부분을 백-엔드 온라인 처리로 수행하는 구조를 제안한다.

2. 백-엔드 온라인 객체 인식 서비스

인공지능 기술을 폭넓게 활용하기 위해 온라인 서비스 형태(AIaaS)로 제공하기 위한 노력이 진행되고 있다 [43]. 기본적으로 전통적인 웹 기반 API 기반 서비스

형태를 취하며 다양한 응용 분야에서 온라인 방식으로 질의-응답을 통한 서비스를 받을 수 있게 진화하고 있다. 딥러닝 모델의 경우, 사전에 학습된 모델을 주로 이러한 온라인 인공지능 서비스에 탑재하고 있으며 기존의 단순 정보검색과 수준에서 문제해결을 위한 딥러닝 모델 기반 서비스 수준으로 올리고 있다.

본 제안 시스템에서 이러한 인공지능 서비스를 자동 레이블링의 기반으로 활용하고 있다. 특히, 레이블링 작업자가 사용하는 작업환경 소프트웨어와 연동하여 클라이언트-서버 모델로 서비스를 제공한다. 레이블링 작업환경은 클라이언트로 다수의 작업자가 손쉽게 본인의 컴퓨터로 참가할 수 있어야 하기 때문에 자동 레이블링 같이 복잡한 라이브러리 및 컴퓨팅 파워를 필요로 하는 부분은 서버에서 부담하는 방식을 취하고 있다. 사용자가 자동 레이블 기능을 선택하면 현재 작업 중인 이미지는 서버로 사전인식의 대상으로 요청된다. 사전 인식은 객체인식 중심으로 다음의 2가지를 제공한다.

1) **박스형 영역 인식** : 사전학습된 YOLO[43] 모델을 통해 입력된 이미지에서 객체를 인식 후, 결과를 반환한다. OpenAPI를 통해 메시지 송수신이 이루어지며, 클라이언트는 인식 결과 메시지를 해석 후, 작업자 화면에 사전인식결과를 표출 후, 작업자가 확인 및 수정 작업을 이어할 수 있도록 한다.

2) **폴리곤형 영역 인식** : 박스형 이상으로 정교하게 이미지 내의 객체영역을 인식하기 위한 것으로, Mask R-CNN[45]을 기본 모델로 한 사전학습 모델을 사용한다. 동일하게 OpenAPI를 통해 수행되며, 박스형과 비교해 폴리곤형으로 표현된 데이터를 반환한다. 인식속도 면에서는 YOLO보다 느리지만, 폴리곤 영역을 작업자가 초기부터 지정하는 것보다 작업량과 속도면에서 우수하다.

상기 2가지 온라인 사전인식서비스는 작업자들이 반복적인 영역 지정을 위한 수고와 시간을 크게 줄여 준다는 측면에서 큰 의미를 갖는다. 특히, 일반적인 이미지 학습데이터 제작량은 점차 증가하고 수십~수백만장에 이를 경우 제작비용과 기간도 역시 크게 늘어날 것이므로 이렇게 학습데이터 생산력을 개선하기 위한

기술 개발과 적용이 중요하다.

다음은 백-엔드 온라인 객체 인식을 통해 자동 레이블링을 수행하기 필수적으로 고려해야 하는 사항이다. 특히, 다중 작업자들의 공동 활용을 위해서는 가능한 다양한 인식모델을 지원해야 하며 빠른 응답을 위해 계산자원을 효과적으로 활용해야 한다.

· **사전인식 클래스명 변환 테이블** : 사전영역 인식 작업과 함께 클래스명을 지정하는 후처리 작업이 필요하다. 이는 개별 작업자별로 다른 클래스군을 목표로 작업할 수 있기 때문에 클라이언트인 작업 소프트웨어 수행하도록 한다. 예를 들어, 사전인식에서 인식된 객체는 객체영역과 객체클래스명으로 반환되는 데, 클래스명은 작업자가 별도로 지정할 수 있게 한다. 만일, 사전인식에서 '트럭'과 같은 클래스명으로 반환되었더라도 작업자는 모두 '차량'으로 지정하길 원한다면 매번 이를 바꾸어 주어야 하는 작업이 필요할 수 있다. 이를 위해 '사전인식 클래스명 변환 테이블'을 설정할 수 있도록 하여, 반환되는 클래스명을 반복적으로 바꾸는 작업을 줄이도록 하고 있다.

· **다중 인식모델 풀과 활성화** : 본 연구에서는 다양한 사전인식모델들을 필요에 따라 제공하기 위한 기능을 개발하였다[46]. 사전 학습된 모델들은 기본적으로 인식모델 풀에서 도커 이미지 형태로 비활성화된 상태로 관리된다. 특히, 본 레이블링 시스템을 다양한 분야에서 공동하기 위해 이러한 모델 관리 기능은 필수적이다. 다만, 이렇게 비활성화되어 있는 모델들은 작업 전에 활성화될 필요가 있다. 활성화를 위해서는 웹 서비스로 오픈하는 작업이 수반된다. 일반적으로 딥러닝 모델들은 추론 시에 학습된 모델 데이터를 CPU 또는 GPU의 메모리로 로딩하는 작업이 필요하기 때문에 매 추론 시마다 이러한 로딩 작업을 반복하는 것은 비효율적이다. 이러한 문제를 해결하기 위해 웹 서비스 개시 전에 메모리에 모델을 로딩하고 상시 추론 수행을 할 수 있도록 하며 이를 통해 전체적인 반응속도를 크게 개선할 수 있었다. YOLO 모델의 경우, 레이블링 시스템에서 이미지를 보내고 나서 응답결과를 받기 위해 평균 1초 내외가 소요되며, Mask R-CNN 모델의 경우는 평균 4초 정도가 소요되었다. 레이블링 작업에서 작업자가 각각 박스 영역 및 폴리곤 영역을 작업하는 것보다 상당

히 빠르며 작업량과 작업시간이 많아짐에 따라 일정한 시간을 보장하는 자동레이블링 기능은 더욱더 효과적이다.

IV. 실험

본 장에서는 제안한 자동 레이블링의 효과를 검증하기 위해 제안 시스템을 활용해 이미지 레이블링 작업 시간을 측정하여, 수작업 대비 개선 효과를 평가한다.

1. 실험 준비

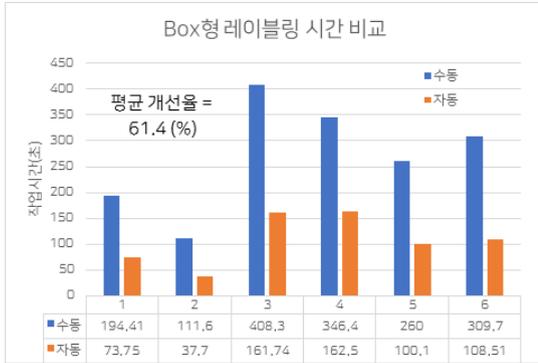
이미지 학습데이터 제작 작업은 작업자의 숙련도 및 이미지 내 객체출현 정도, 객체 형태의 복잡도에 따라 크게 달라질 수 있기 때문에 본 평가에서는 임의의 영상을 중심으로 제한된 실험으로 수행하였다.

· **실험데이터셋** : 본 실험에서는 박스형과 폴리곤형의 레이블링 작업에서 있어 사전인식을 통한 레이블링 속도 개선율을 평가한다. 일반적으로 폴리곤형의 레이블링에 더 많은 작업시간이 소요되며 사전인식결과를 활용해 레이블링 시간을 더 효과적으로 단축할 수 있을 것으로 예상된다. 특히, 이미지 내에 포함된 객체의 개수에 따라 레이블링 시간이 크게 달라지므로 실험데이터셋은 이러한 차이를 살펴볼 수 있도록 구성할 필요가 있다. 본 실험에서는 [그림 2]에서 보이는 바와 같이 10장의 이미지로 각각의 이미지를 제안 레이블링 시스템을 사용하여 작업을 수행한다. 이 그림에는 추가적으로 Mask R-CNN 모델을 활용하여 사전인식한 결과를 보여주고 있으며 작업자는 이를 수정하여 최종 레이블링 작업을 마칠 수 있게 된다. 실험에는 객체 개수를 2개부터 10개까지 포함하는 다양한 이미지를 준비하여 대량의 이미지에 대한 개선율 측정보다 소수의 사례를 중심으로 자동 레이블링의 효과를 상세히 살펴 본다.

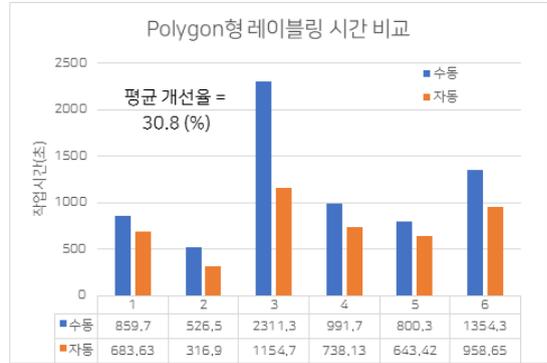
· **작업환경** : 실험에 참가하는 레이블링 작업자는 본 연구를 통해 준비된 레이블링 시스템을 활용하여 작업을 수행하게 된다. 작업자에게는 10장의 이미지가 순차적으로 주어지며 수동 및 자동 방식에 의해 레이블링을 수행하게 된다. 작업들에는 박스 및 폴리곤 형태로 이미지 상에서 영역을 지정하고 각 객체 영역에 해당하는 클래스를 선택하는 방식으로 진행하게 된다.



그림 2. 실험용 이미지: Mask R-CNN 기반 사전인식 결과 활용

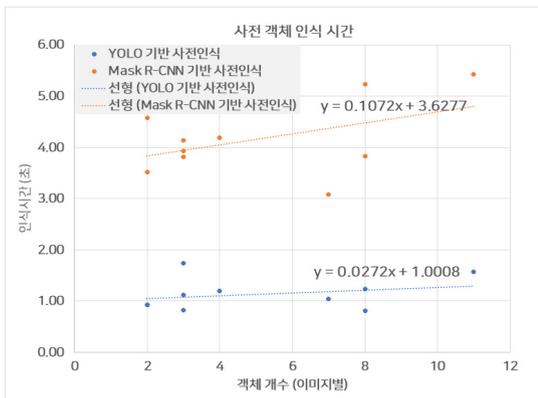


(1) 박스형 레이블링 작업시간

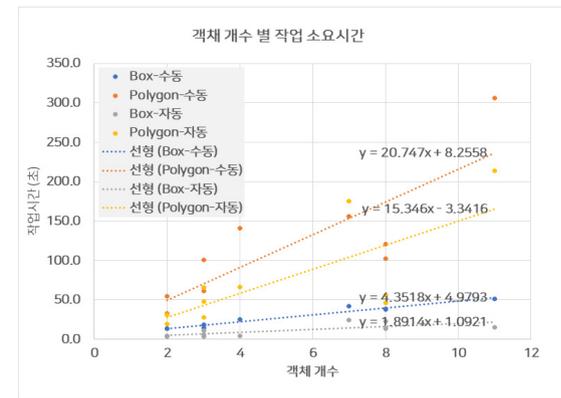


(2) 폴리곤형 레이블링 작업시간

그림 3. 레이블링 형태별(Box/Polygon) 수동/자동 작업시간 비교



(1) 사전객체 인식시간



(2) 레이블링 작업시간

그림 4. 객체 개수 별 소요시간 비교

실험에 소요되는 부가적인 이미지 로딩 및 결과 출력 시간은 레이블링 시간에 포함되지 않는다. 타겟으로 하는 객체의 클래스는 car, bus, truck, person, elephant, zebra, bear, cat, dog, sheep의 총 10종으로 시스템에 미리 클래스명이 등록되어 있다. 이러한 준비작업은 한 번만 등록하고 반복적으로 사용되므로 레이블링 시간에 역시 포함되지 않는다.

· **실험 참가자** : 실험에는 [표 3]과 같이 30~40대 연령의 IT 관련 분야 전공 연구자들(남성) 6명이 참가하였다. 실험참가에 대한 동일한 조건을 유지하기 위해 참가자들은 이러한 레이블링 작업에 대해 미경험자들로, 실험 전에 작업 목표 및 작업환경에 대한 충분한 설명과 함께 작업툴을 사용해 볼 수 있는 시간이 미리 주어졌다. 또한, 참가 작업자들은 작업 중에 오로지 레이블링 작업에만 집중하도록 하며 개인별로 이미지 내에서 영역에 대한 지정 방식에는 특별히 제약을 주지 않고 자유롭게 하되 최대한 객체를 정확히 커버링하도록 요구하였다. 본 실험에서는 작업자들의 연령과 작업량의 직접적인 상관관계는 나타나고 있지 않으며 향후 연구로 남겨 둔다.

표 3. 작업자 연령 정보

작업자ID	1	2	3	4	5	6
연령	36	41	43	46	38	34

2. 실험 수행 내용 및 결과 분석

각 실험자에게는 [그림 2]에 제시된 10장의 이미지에 대해 수동과 자동 레이블링 방식으로 작업을 하도록 요구하였다. 특히, 각각에 실험에 대해 박스형과 폴리곤형의 영역지정을 수행하게 하였다.

1) 수동 레이블링 : 작업자가 객체를 각 이미지부터 처음부터 찾고, 영역을 지정하고 나서 해당하는 클래스를 선택해야 하는 과정을 반복하게 된다. 다만, 작업자들에게 별도로 레이블링해야 하는 객체를 제시하여 반드시 레이블링해야 하는 객체 개수를 유지하도록 하였다. 수동방식이다 보니 작업자별로 객체영역을 지정하는 방법이 달라 시간차가 발생하게 된다. 특히, 박스형보다 폴리곤형의 영역 지정에 시간이 많이 소요되며,

이미지 내의 객체 수도 비례한다. 결과적으로 객체수가 증가함에 따라 작업자들에게 상당한 부담을 요하기 때문에 피로도가 높은 작업이었다.

2) 자동 레이블링 : 시스템을 통해 각 이미지를 여는 순간부터 자동적으로 백-엔드 사전 레이블링을 요청하여 결과를 받아 사용자 작업화면에 표출하도록 되어 있다. 사전 객체 인식에 소요되는 시간은 [그림 4](1)과 같으며, YOLO 모델을 활용하는 박스형의 경우 개체수 변화에 따라 미미한 변화를 보이며 평균 1.14초 내에 작업자에게 결과를 반환하였다. Mask R-CNN 모델을 활용하는 폴리곤형의 경우, 객체 개수 증가에 따라 점진적으로 증가하긴 하지만, 본 실험에서의 개체 수 범위 내에서는 평균 4.17초 내에 결과를 반환하였다. 작업자가 각 작업을 수동방식으로 이러한 사전인식 시간 내에 끝낼 수 있다면 자동방식이 효과적이라고 할 수 없겠지만, 사전인식 시간이 전반적으로 수 초내에 끝나며 작업자는 바로 이를 활용하여 수정/추가 작업을 통해 작업을 마칠 수 있어 자동 방식이 더 효과적이라고 할 수 있다. 특히, 본 연구에서는 작업자가 사용하는 레이블링 시스템에서 다수의 이미지를 동시에 작업할 경우, 백-그라운드에서 다음 작업이미지에 대한 자동 레이블링을 요청하여 사전에 수행해 놓기 때문에, 작업자는 다음 작업이미지에서 대기할 필요없이 작업을 개시할 수 있다. 따라서, 1~4초 정도 소요되는 자동 레이블링 처리 시간은 사실상 무시할 수 있게 된다.

· **수동 vs. 자동 레이블링 시간 분석** : [그림 3]에는 박스형과 폴리곤형에 대한 작업시간 분석결과를 그래프로 표시하고 있다. [그림 3](1)의 박스형 레이블링의 결과 6명의 작업자마다 작업시간이 달라지지만 자동방식을 활용하는 경우, 시간이 모두 단축되었으며 평균개선율(=(수동방식 소요시간 - 자동방식 소요시간)/(수동방식 소요시간) * 100(%))이 61.4%까지 나타났다. 즉, 작업시간을 약 1/3 수준으로 단축할 수 있게 되었다. 폴리곤형 레이블링의 경우 [그림 3](2)와 같이 마찬가지로 모두 작업시간을 단축할 수 있었으며 평균개선율이 30.8%로 약 2/3 수준으로 시간을 단축할 수 있었다. 폴리곤형의 경우, 예상대로 작업자에게 요구되는 사전 인식결과의 수정에 있어서 박스형보다 다소 복잡하며,

폴리곤을 구성하는 다수의 점(Point)들을 이동하는 방식으로 작업이 필요했다. 인터페이스상의 제약으로, 처음부터 수동으로 지정하는 것과 비교해 원하는 영역을 정확하게 표현하기 힘든 제약사항이 있었지만, 최대한 사전인식 결과를 수정하는 수준에서 작업을 수행하도록 하였다. 결과적으로, 박스형과 폴리곤형 레이블링 모두에서 30~60%의 작업시간을 단축할 수 있었으며, 특히 레이블링을 수동방식으로 처음부터 수행하는 것이 아닌 사전인식 결과를 수정하는 방식이라 작업자의 작업부담은 크게 개선되었다.

· **레이블링 작업시간의 예측** : 본 연구에서 각 이미지에 대한 레이블링 시간에 영향을 미치는 변수로는 1) 이미지 내의 객체 수와 2) 레이블링 방식(박스형 또는 폴리곤형)이다. 이 2개의 변수에 따른 작업시간은 [그림 4](2)와 같으며 '선형(Polygon-수동)'으로 나타낸 선은 수동으로 폴리곤형의 레이블링을 수행할 때 요구되는 시간에 대한 선형회기 결과를 나타내며, $y=20.746x+8.2558$ 의 식으로 표현된다. 즉, 폴리곤형으로 객체를 한 개씩 지정할 때마다 약 20초 정도씩 추가적으로 소요됨을 의미한다. 마찬가지로, 수동과 자동으로 박스형, 폴리곤형의 레이블링에 대한 소요시간 예측결과를 나타낸다. 이를 바탕으로 평균적으로 6개의 객체를 갖는 1만개의 이미지에 대해 소요되는 시간을 [그림 4](1)의 회기식들에 따라 예측해 보면 [표 4]과 같다. 여기서는 각 작업자는 1일 8시간씩 작업하는 것으로 가정하며, 작업자들의 휴식시간과 집중력 저하를 고려하지 않는 단순 산술결과이다. 이러한 방식으로 대량의 이미지 학습데이터 제작에 소요되는 시간을 예측하고, 다수의 작업자를 활용하여 보다 계획적으로 학습데이터를 생산할 수 있다.

표 4. 레이블링 작업량 예측 결과

1만장 이미지(평균 6개 객체 포함)에 대한 레이블링 작업량, 1인 8시간/1일 수행		
	박스형	폴리곤형
수동	10.79일	46.08일
자동	4.32일	33.13일

V. 제안 실험의 한계 및 향후 계획 상세

본 연구에서는 이미지 학습데이터 제작에 있어 사전 인식 결과를 활용한 자동레이블링의 효과를 검증하기 위한 실험을 다음과 같이 한정된 범위에서 수행하였으며, 보다 개선된 자동레이블링 결과를 얻기 위해 향후 연구 및 시스템 개발을 이어 나갈 계획이다.

· **실험 데이터 개수와 객체 다양성** : 본 실험에서는 [그림 2]의 10장의 실험 이미지를 선정하여 객체 개수 증가 따른 효과를 집중적으로 분석하였다. 실제적인 대량의 이미지 레이블링 작업에서는 더 많은 데이터로부터 더 복잡하고 다양한 객체형태를 갖는 경우가 발생할 수 있으며 이에 대해 향후 연구를 통해 더 많은 사례에 대한 개선방안을 연구할 계획이다.

· **레이블링 작업자의 작업 스타일** : 일반적으로 이미지 레이블링에서 영역지정범위는 작업자별로 조금씩 다를 수 밖에 없는 한계로 본 실험에서는 이에 대한 제약을 크게 두지 않고, 반드시 포함되어야 하는 객체 수만을 지정하였다. 또한, 참가자에 따라서는 폴리곤 방식에 있어 하나의 객체를 지정하기 위해 다수의 점들로 지정된 영역을 표현하기 위해 많은 시간을 소모하기도 하여 작업자별로 서로 다른 양상을 보일 수 있다. 이러한 개인차 및 이에 대한 개선방안에 대해서도 향후 연구를 통해 안정적인 품질의 레이블링을 위한 기법에 대한 연구를 진행할 계획이다.

· **사전 레이블링 모델 성능** : 사전인식을 위해 본 실험에서는 YOLO 모델과 Mask R-CNN 모델을 COCO데이터셋[6]으로 훈련시킨 사전학습모델을 사용하였다. 특히, 본 연구는 이들 모델로 발견가능한 10종의 객체를 미리 선정하여 진행하여 객체영역 지정 성능 중심으로 그 효과를 분석하였다. 하지만, 실제적으로는 이러한 모델로 제대로 발견되지 못하는 경우가 발생할 것이며 작업자가 이를 수정해도 사전학습모델이 이를 다시 활용하여 배울 수 있는 방법이 마련되어 있지 않다. 이러한 Active Learning[47][48]과 같은 방식으로 작업자의 중간 작업 산출물을 바탕으로 사전인식 성능을 개선시켜 나아가는 방법을 생각할 수 있으며 이를 적용하여, 작업하면 할수록 수정해야 하는 일의 양을 줄일 수 있을 것으로 예상된다. 이러한 훈련방법을 적

용한 자동레이블링 방법에 대해서도 향후 연구를 통해 구체적인 적용 모델 설계와 성능 평가를 수행할 계획이다.

· **레이블링 작업 도구의 사용성** : 본 실험에서는 사전 인식모델 결과를 박스형 또는 폴리곤형의 영역으로 이미지와 함께 작업자에게 제시하고 수정을 할 수 있도록 하고 있다. 실험참가자들에게 작업S/W에 익숙할 시간을 충분히 주었지만 영역 수정에 있어 여전히 작업툴 사용이 쉽지 않는 문제가 있었다. 특히, 폴리곤형의 영역 수정에서 점들의 영역을 고치기 위해 각 점을 선택하는 경우 GUI상에서 이를 잡기 어렵거나, 또는 점을 추가하여 폴리곤 영역을 넓히는 방법이 마련되어 있지 않은 문제가 있었다. 즉, 자동레이블링 결과를 보다 손쉽게 수정할 수 있게 GUI를 구현해야만 실제적으로 사전인식결과 활용의 효과를 얻을 수 있어 이에 대한 개선이 필요함을 알 수 있었으며, 향후 레이블링 시스템 기능 보완 및 추가 개발을 통해 이러한 문제점을 해결할 예정이다.

VI. 결론 및 향후 계획

본 논문에서는 영상 학습데이터 제작에 필요한 작업자의 레이블링 작업을 줄이기 위해 전처리로서 객체탐지를 통해 사전 레이블링을 수행하고, 작업자에게 이를 확인 및 수정하게 하는 방식의 시스템을 제안하고 있다. 특히, 다양한 분야에서의 학습데이터 제작 니즈에 대응하기 위해 백-엔드 온라인 인식 서비스와 연동한 레이블링 기법을 통해 지속적으로 인식모델을 확대 및 업데이트할 수 있는 이점을 가지고 있다. 실험용 이미지셋을 통한 수동과 자동 레이블링 비교 실험결과에서도 자동 레이블링이 작업속도 면에서도 크게 효과적이며, 특히 박스형과 폴리곤형의 레이블링 방식 모두에서 작업시간을 단축할 수 있었다.

향후 계획으로는 앞서 5장에서 언급한 바와 같이 실험상에서의 한계 극복과 시스템 개선을 위한 추가적인 연구를 진행할 예정이다. 특히, 레이블링 작업 중에서도 지속적으로 누적되는 데이터를 활용하여 백-엔드 인식 모델을 재학습 및 업데이트를 통해 제공하기 위한 기술

과 서비스 시스템을 개발할 예정이다. 또한, 작업자별 레이블링 결과의 품질 차이를 줄이기 위해 레이블링 작업 워크플로우 관리 및 자동 품질 평가 기술에 관한 연구개발을 수행할 예정이다.

참고 문헌

- [1] Cognilytica, "Data Engineering, Preparation, and Labeling for AI 2019," <https://www.cognilytica.com/2019/03/06/report-data-engineering-preparation-and-labeling-for-ai-2019/>.
- [2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *IEEE*, Vol.86, No.11, pp.2278-2324, 1998.
- [3] Alex Krizhevsky, Learning Multiple Layers of Features from Tiny Images, <https://www.cs.toronto.edu/~kriz/cifar.html>, 2009.
- [4] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, Schiele, and Bernt, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014. <http://human-pose.mpi-inf.mpg.de/>
- [5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision* 88, pp.303-338, 2010.
- [6] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick, "Microsoft COCO: Common Objects in Context," <https://cocodataset.org>, 2014.
- [7] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fe, "Imagenet: A large-scale hierarchical image database," In: 2009 IEEE conference on computer vision and pattern recognition, pp. 248-55, 2009, <http://www.image-net.org/>
- [8] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun, "Vision meets robotics: T

- he KITTI dataset,” *Int. J. Robotics Res.*, Vol.32, No.11, pp.1231-1237, 2013, <http://www.cvlibs.net/datasets/kitti/>
- [9] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The Cityscapes Dataset for Semantic Urban Scene Understanding,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, <https://www.cityscapes-dataset.com/>
- [10] IMDb Datasets, <https://www.imdb.com/interfaces>
- [11] Google’s Open Images V6 + Extensions, <https://storage.googleapis.com/openimages/web/index.html>
- [12] Lucas Smaira, João Carreira, Eric Noland, Ellen Clancy, Amy Wu, and Andrew Zisserman, *A Short Note on the Kinetics-700-2020*, Human Action Dataset, CoRR abs/2010.10864, 2020, <https://deepmind.com/research/open-source/kinetics>
- [13] The 20BN-jester Dataset V1, <https://20bn.com/datasets/jester>
- [14] Kaggle: Your Machine Learning and Data Science Community, <https://www.kaggle.com/>
- [15] AI Hub, <https://aihub.or.kr/>
- [16] H. Emami, M.M. Aliabadi, M. Dong, and R.Chinnam, “SPA-GAN: Spatial Attention GAN for Image-to-Image Translation,” *IEEE Trans. on Multimedia*, Vol.23, pp.391-401, 2021.
- [17] H. Ko, D. Y. Lee, S. Cho, and A. C. Bovik, “Quality Prediction on Deep Generative Images,” *IEEE Trans. on Image Processing*, Vol.29, pp.5964-5979, 2020.
- [18] Gregory Griffin, Alex Holub, and Pietro Perona, *Caltech-256 Object Category Dataset*, California Institute of Technology, 2007, <https://www.kaggle.com/jessicali9530/caltech256>
- [19] Google’s Conceptual Captions, <https://ai.google.com/research/ConceptualCaptions/>
- [20] LabelMe, <https://github.com/wkentaro/labelme>
- [21] MS VoTT, <https://github.com/microsoft/VoTT>
- [22] Make-Sense, <https://www.makesense.ai/>
- [23] SuperAnnotate, <https://superannotate.com/>
- [24] VIA(VGG), <https://www.robots.ox.ac.uk/~vgg/software/via/>
- [25] Supervise.ly, <https://supervise.ly/>
- [26] Labelbox, <https://labelbox.com/>
- [27] Hasty.ai, <https://hasty.ai/>
- [28] CVAT, <https://github.com/openvinotoolkit/cvat>
- [29] Darwin, <https://www.v7labs.com/darwin>
- [30] Heartex, <https://heartex.com/>
- [31] Scalabel, <https://www.scalabel.ai/>
- [32] Segments.ai, <https://segments.ai/>
- [33] LabelIMG, <https://github.com/tzutalin/labelimg>
- [34] ImageTagger, <https://github.com/bit-bots/image-tagger>
- [35] DarkLabel, <https://github.com/darkpgmr/DarkLabel>
- [36] EVA, <https://github.com/Ericsson/eva>
- [37] LOST, <https://github.com/l3p-cv/lost>
- [38] Ybat, <https://github.com/drainingsun/ybat>
- [39] MuViLab, <https://github.com/ale152/muvilab>
- [40] Turkey, <https://github.com/yanfengliu/turkey>
- [41] Point Cloud Annotation Tool, <https://github.com/springzfx/point-cloud-annotation-tool>
- [42] DeepLabel, <https://github.com/jveitchmichaelis/deeplabel>
- [43] 이용 외, “인공지능 서비스(AIaaS) 기술 동향과 활성화 방안,” *정보과학회지*, 제38권, 제8호, pp.49-57, 2020.
- [44] YOLO: Real-Time Object Detection, <https://pjreddie.com/darknet/yolo/>
- [45] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” 2017 IEEE International Conference on Computer Vision (ICCV), pp.2980-2988, 2017.
- [46] 장래영, 이용, 박민우, 이상환, “OpenFaaS 기반 AI 분석 서비스 시스템 구축,” *한국콘텐츠학회논문지*, 제20권, 제7호, pp.97-106, 2020.
- [47] B. Settles, *Active Learning Literature Survey*, University of Wisconsin—Madison, 2009.
- [48] R. Nowak and S. Hanneke, “Active Learning: From Theory to Practice,” *ICML2019 Tutorial*, 2019.

저 자 소 개

이 용(Ryong Lee)

정회원



- 2003년 : 일본교토대학교 정보학과 (공학박사)
- 2013년 9월 ~ 현재 : 한국과학기술정보연구원 연구데이터공유센터 책임연구원
- 2021년 3월 ~ 현재 : 과학기술연합대학원대학교 교수

〈관심분야〉 : 인공지능, 시각지능, 사물인터넷, 공간데이터 등

장 래 영(Rae-Young Jang)

정회원



- 2018년 8월 : 한남대학교 컴퓨터공학과(공학박사)
- 2019년 9월 ~ 현재 : 한국과학기술정보연구원 연구데이터공유센터 박사후연구원

〈관심분야〉 : DevOps, Docker, Kubernetes, 인공지능 등

박 민 우(Min-Woo Park)

정회원



- 2004년 : 충남대학교 컴퓨터과학과 (공학석사)
- 1996년 8월 ~ 현재 : 한국과학기술정보연구원 개방형데이터융합연구단 책임연구원

〈관심분야〉 : 시스템 아키텍처, 정보보안, 인공지능 등

이 건 우(Gunwoo Lee)

정회원

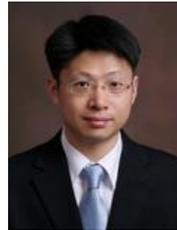


- 2020년 : 한국과학기술원 전산학과 (공학박사)
- 2020년 12월 ~ 현재 : 한국과학기술정보연구원 연구데이터공유센터 선임연구원

〈관심분야〉 : 인공지능, 네트워크, 지능형 시스템 등

최 명 식 (Myung-Seok Choi)

정회원



- 2005년 : 한국과학기술원 전산학과 (공학박사)
- 2005년 ~ 현재 : 한국과학기술정보연구원 연구데이터공유센터 책임연구원

〈관심분야〉 : 오픈사이언스, 연구데이터관리, 인공지능 등