

객체 분할 기법을 활용한 자동 라벨링 구축

문준휘¹ · 박성현¹ · 최지영¹ · 신원선² · 정회경^{1*}

¹배재대학교 ²vision21테크

Auto Labelling System using Object Segmentation Technology

Jun-hwi Moon¹ · Seong-hyeon Park¹ · Jiyoung Choi¹ · Wonsun Shin² · Heokyung Jung^{1*}

¹Paichai University · ²vision21tech

E-mail : answngn123@naver.com / 1861027@pcu.ac.kr / aurily2017@naver.com /
vision21tech@naver.com / hkjung@pcu.ac.kr

요 약

객체 분할 분야의 딥러닝 기반 컴퓨터 비전 응용들은 성능을 향상하기 위하여 STOA 기법들이 사전 학습하여 배포한 하이퍼파라미터와 모델을 통해 학습하는 전이학습 방법을 사용한다. 이 과정에서 사용되는 커스텀 데이터 셋들은 Ground Truth 정보를 생성하기 위한 라벨링 작업에서 시간이나 라벨러등의 많은 자원을 필요로 한다. 본 고에서는 딥러닝 신경망에서 사용되는 커스텀 데이터 셋 구축을 위하여 시간이나 라벨러등의 자원을 적게 사용할 수 있도록 객체 분할 기법을 활용한 자동 라벨링 구축 방법을 제시한다.

ABSTRACT

Deep learning-based computer vision applications in the field of object segmentation take a transfer learning method using hyperparameters and models pretrained and distributed by STOA techniques to improve performance. Custom datasets used in this process require a lot of resources, such as time and labeling, in labeling tasks to generate Ground Truth information. In this paper, we present an automatic labeling construction method using object segmentation techniques so that resources such as time and labeling can be used less to build custom datasets used in deep learning neural networks.

키워드

Deep learning, Object Segmentation, Custom Dataset, Auto Labelling

1. 서 론

컴퓨터 비전은 이미지와 비디오로부터 정보를 구분해 내는 기술로써 다양한 분야에서 활용되고 있으며 특히 인공지능 기술이 접목되어 빠른 발전을 이루고 있다[1]. 인공지능 기반 컴퓨터 비전은 딥러닝을 적용한 이미지 인식 기법이 이미지 인식 대회인 ImageNet Large Scale Visual Recognition Challenge(ILSVRC)에서 2012년 우승하였고 Top-5에 대한 인식의 정확도가 사람의 인식 수준보다 더 나은 98.7%까지 2020년도에 달성하는 등 실제 응용에 사용할 수 있게 되어 적용 분야를 점점 넓히

고 있다[2, 3].

인공지능 기반 컴퓨터 비전 응용을 만들기 위해서는 응용에 적합한 모델을 구축하고 모델을 학습시킨 후 배포하여 새로운 데이터에 적용하여 예측하는 일반적인 과정을 수행한다. 모델의 학습에 사용되는 데이터는 이미지 인식 대회에서 성능을 비교하기 위해 사용되는 범용 목적의 데이터를 사용하는 것보다 응용에 적합한 커스텀 데이터를 사용하여 모델을 학습시키는 것이 더 나은 성능을 보일 수 있다.

커스텀 데이터 셋을 구축하기 위해 1) 데이터를 수집하고 2) 수집한 데이터에 대하여 라벨링을 진행하여 사용하는데 이 작업은 주로 사람이 진행한

* corresponding author

다. 라벨링 과정은 시간이 오래 필요한 작업이며 라벨링 작업자에 따라 다른 형태로 라벨링이 이루어질 수 있고 작업자의 피로도에 따라 생성되는 라벨링 데이터의 품질이 결정될 수 있다[3, 4]. 라벨링 작업자에 의한 라벨링 데이터의 품질 저하를 최소화하고 데이터 셋의 라벨링 정보가 표준화된 형태로 생성될 것과 데이터의 양이 대량일 때 빠르게 진행되도록 하는 것이 양질의 커스텀 데이터 셋을 구축하는 데 도움이 될 것이다.

본 논문에서는 커스텀 데이터 셋을 구축하기 위해 표준화된 형태의 라벨링 정보를 생성하고 비교적 짧은 시간 동안 진행되도록 하기 위한 자동 라벨링 시스템을 제안한다. 2장에서는 이미지 인식에 사용되는 일반적 데이터 셋의 형태와 커스텀 데이터 셋에서 구축되어야 할 정보의 형태에 관하여 기술한다. 3장에서는 커스텀 데이터 셋을 구축하기 위한 자동 라벨링 시스템의 구조를 제안하고 4장에서는 향후 연구 방향에 대해 제안하고자 한다.

II. 본 론

2.1 선행연구

객체 분할 기법을 사용한 자동 라벨링 시스템은 Mask R-CNN 기법을 기반으로 구현한 시스템이 있다[3, 5]. Mask R-CNN 기반 객체 분할 기법은 인식 성능은 우수하지만 프레임 처리 속도는 느리다. 따라서 작업 시간이 많이 요구되어 작업자를 대신하는 자동 라벨링의 이점을 감소시킨다.

2.2 객체 탐지용 공개 데이터 셋

제안하는 라벨링 작업은 객체 분할 기법을 활용하는 방법이다. 객체 분할은 객체 탐지로부터 발전된 기술로서 공개된 데이터 셋인 MS COCO 데이터 셋이나 PASCAL VOC 데이터 셋을 많이 사용하고 있다[4].

표 1. PASCAL VOC vs MS COCO

	Target Field	Purpose	Scale	Data Format	Label Structure	Producer
PASCALVOC	Image Recognition	Classification/ Detection	20classes, 9.963Images	XML	Image-Label-image-Region-Label	Mark Everingham, et. al
MSCOCO	Image Recognition	Classification/ Detection/ Caption	100classes2.5millioninstances	JSON	Image-Label-image-Region-Label	COCO Consortium

1) MS COCO 데이터 셋[6]

2015년부터 2020년까지 대회를 진행하고 있는 MS COCO 데이터 셋은 인공지능 모델의 성능을 향상시키기 위한 목적으로 만들어졌다. 학습데이터들이 하나의 이미지에 하나의 객체가 존재하는 형태를 일반화시킨 것으로 객체 감지, 세그멘테이션 분야의 학습 데이터로 많이 사용된다. 하나의 이미지에 평균 7개 이상의 객체가 존재하는 형태의 이미지들로 구성되어 있으며 이미지내 객체들에 대해 82개의 카테고리로 분류하고 JSON 형태로 이미지 라벨링 정보를 저장한다.

2) PASCAL VOC[7]

객체 분류를 위한 목적으로 만들어진 데이터 셋으로 XML 형태로 라벨링 정보를 저장한다. 객체 분류, 객체 탐지, 세분화 평가 알고리즘을 구축하거나 평가하는 데 자주 사용된다. 20여개의 카테고리로 분류하고 라벨링되는 데이터 셋들은 Annotations, ImageSet, JPEGImages, SegmentationClass, SegmentationObject으로 폴더 구조를 가진다.

2.3 커스텀 데이터 셋 구축

객체 감지 모델을 만들기 위해 모델을 학습시키는 방법도 중요하지만, 목표하고자 하는 응용에 적합한 데이터 셋도 필요하다. 공개데이터 셋인 MS COCO 데이터 셋이나 PASCAL VOC 데이터 셋을 활용하는 것은 타겟으로 삼고 있는 객체를 원활하게 찾지 못할 수도 있다[9]. 따라서, 목표에 적합한 커스텀 데이터를 구축해야 하지만 품질이 좋은 대량의 이미지 셋을 찾기 힘들고, 수집한 데이터 셋에 대해 라벨링을 하는 과정은 많은 시간과 노력이 요구된다.

그러나 응용에 적합한 분류기를 만드는 것은 객체 검출이나 분류 기능을 수행하는 모델 설계도 중요하고, 모델 학습을 위한 학습데이터도 매우 중요하다[4].

III. 제안하는 시스템

제안하는 시스템은 Mask R-CNN 기반 객체 분할 기법과 성능은 비슷하지만, 처리 속도는 더 빠른 YOLACT를 사용하여 구축한다[8]. 제안하는 시스템은 데이터 수집부, 데이터 저장부, 객체 세그멘터부로 구성된다. 데이터 수집 부에서는 커스텀 데이터 셋에 사용될 원시 이미지 데이터를 인터넷으로부터 수집하는 크롤링 기능, 비디오 영상으로부터 프레임별 이미지를 추출하는 프레임 추출 기능으로 구성되며 원시 이미지 데이터가 이미 존재하는 경우 구동하지 않을 수 있다. 데이터 저장부는 수집된 원시 이미지를 클래스에 따라 이미지를 분류하여 저장하는 기능을 수행한다. 객체 세그멘터부에서는 데이터 저장부를 통해 저장된 이미지 데이터에 대해 라벨링 데이터 생성을 담당한다. 이를 위하여 라벨링 데이터의 형식을 PASCAL VOC 형태와 MS COCO 형태 중 선택할 수 있도록 제공하고, 객체 분할 기능을 구동하여 라벨링 데이터를 생성하여 저장하는 기능을 수행한다.

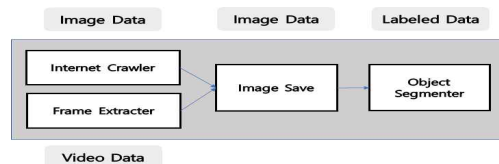


그림 1. 자동라벨링 Processing Flow Chart

객체 분할 기능을 적용한 자동 라벨링 시스템은 라벨링 처리를 한 학습데이터와 사전학습된 모델을 사용하여 전이 학습을 진행한다. 이 과정을 통해 생성된 모델을 사용하여 라벨링 처리가 되어 있지 않은 원시 이미지 데이터에 대한 객체 분할을 진행한다. 객체 분할을 진행하기 전 기능의 실행을 통해 자동 생성될 라벨링 데이터의 형식을 선택하도록 하고 기능을 수행하여 라벨링된 인공지능용 컴퓨터 비전 모델 학습용 이미지 데이터를 완성한다[6, 7].

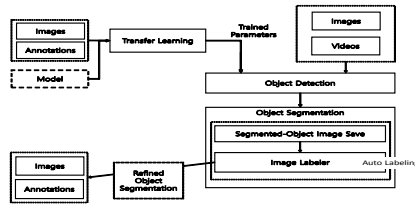


그림 2. 제안하는 자동 라벨링 시스템 구조

IV. 결 론

인공지능 서비스를 제공하려면 서비스에 적합한 모델이 필요하고 모델을 제공하기 위해서는 모델의 학습이 진행되어야 한다. 모델의 학습을 위해 요구되는 응용 적합 커스텀 데이터 셋을 구축하기 위하여 본 논문에서는 커스텀 데이터 셋을 구축하는 자동 라벨링 방법에 대하여 제안하였다.

인터넷의 이미지 데이터를 수집하고 비디오 영상으로부터 프레임을 추출하여 이미지 데이터를 구축하였다. 수집한 데이터 셋에 대해 학습을 위한 클래스 구조로 저장하고 객체 분할 기능을 활용하여 자동 라벨링이 수행되도록 기능을 구축하였다. 웹카메라의 입력 영상을 사용하는 경우에도 라벨링이 가능하도록 처리 속도에 중점을 두고 자동 라벨링 시스템을 제안하였다.

본 시스템에서 기반으로 한 YOLACT는 Mask R-CNN 기반의 객체 분할 기법에 비해 객체 분할 성능이 다소 떨어지지만 처리 속도면에서 월등히 빠르다[6]. 따라서 처리 속도의 이점을 유지하면서 성능을 개선하는 방안에 대해 시스템 구조적인 측면과 이미지를 생성하고 연산하는 방식을 수정해 보는 등 연구를 진행하고 있다.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW) (No. 2019-0-01838) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

References

- [1] J. Lee, "Artificial Intelligence Image Recognition Technology Trend," *TTA Journal*, vol. 187, TTA, pp. 44-51, Jan. 2020.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *NIPS*, 2012.
- [3] Q. Xie, M. Luong, E. Hovy, and Q. V. Le, "Self-training with Noisy Student improves ImageNet classification", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10687-10698, Jun. 2020.
- [4] R. Lee, R. Jang, M. Park, G. Lee, and M.-S. Choi, "An Auto-Labeling based Smart Image Annotation System," *The Journal of the Korea Contents Association*, vol. 21, no. 6, pp. 701-715, Jun. 2021.
- [5] H. Jo, D. Kim, and J.-B. Song, "Automatic Dataset Generation of Object Detection and Instance Segmentation using Mask R-CNN," *Journal of Korea Robotics Society*, vol. 14, no. 1, pp. 31-39, Mar. 2019.
- [6] T. Lin, M. Maire, S. Belongie, J. Hayes, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *European Conference on Computer Vision*, Springer, pp. 740-755, 2014.
- [7] 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*, vol. 88, no. 2, pp. 303-338, 2010.
- [8] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time Instance Segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9157-9166, Oct. 2019.
- [9] H. Lee, and H. Jung, "Auto Labeling Restaurant Notices of Delivery Applications using Text Classification Model", in *Proceedings of Korea Computer Congress 2022*, vol. 26, no. 2, pp. 1967-1969, 2022.