

https://doi.org/10.7236/JIIBC.2020.20.1.219
JIIBC 2020-1-31

YOLO 네트워크를 활용한 전이학습 기반 객체 탐지 알고리즘

Transfer Learning-based Object Detection Algorithm Using YOLO Network

이동규*, 선영규*, 김수현*, 심이삭*, 이계산**, 송명남***, 김진영****

Donggu Lee*, Young-Ghyu Sun*, Soo-Hyun Kim*, Issac Sim*,
Kye-San Lee**, Myoung-Nam Song***, Jin-Young Kim****

요약 딥 러닝 기반 객체 탐지 및 영상처리 분야에서 모델의 인식률과 정확도를 보장하기 위해 다량의 데이터 확보는 필수적이다. 본 논문에서는 학습데이터가 적은 경우에도 인공지능 모델의 높은 성능을 도출하기 위해 전이학습 기반 객체 탐지 알고리즘을 제안한다. 본 논문에서는 객체탐지를 위해 사전 학습된 Resnet-50 네트워크와 YOLO(You Only Look Once) 네트워크를 결합한 전이학습 네트워크를 구성하였다. 구성된 전이학습 네트워크는 Leeds Sports Pose 데이터셋의 일부를 활용하여 이미지에서 가장 넓은 영역을 차지하고 있는 사람을 탐지하는 네트워크로 학습을 진행하였다. 실험결과는 탐지율 84%, 탐지 정확도 97%를 기록하였다.

Abstract To guarantee AI model's prominent recognition rate and recognition precision, obtaining the large number of data is essential. In this paper, we propose transfer learning-based object detection algorithm for maintaining outstanding performance even when the volume of training data is small. Also, we proposed a transfer learning network combining Resnet-50 and YOLO(You Only Look Once) network. The transfer learning network uses the Leeds Sports Pose dataset to train the network that detects the person who occupies the largest part of each images. Simulation results yield to detection rate as 84% and detection precision as 97%.

Key Words : Object Detection, Transfer Learning, Deep Learning, Image Processing

1. 서론

전이학습은 미리 학습된 인공신경망을 활용하여 처음에 학습된 데이터와 유사한 데이터의 특징을 학습하여

사용하는 기법으로, 처음에 이용한 학습데이터 보다 더 적은 수의 데이터로 새로운 데이터의 특징을 학습한다^[1]. 이러한 특징을 이용하여 다양한 분야에서 활용한 연구들이 뛰어난 성능을 보이고 있다^{[3]-[5]}.

*준회원, 광운대학교 전자융합공학과

**정회원, 경희대학교 전자공학과

***준회원, (주)현대엠아이비인터내셔널

****정회원, 광운대학교 전자융합공학과

접수일자: 2019년 11월 25일, 수정완료: 2020년 1월 5일

게재확정일자: 2020년 2월 7일

Received: 25 November, 2019 / Revised: 5 January, 2020 / Accepted: 7 February, 2020

****Corresponding Author: jinyoung@kw.ac.kr

Dept. of Electronic Convergence Engineering, Kwangwoon Univ., Seoul, Korea

[3]의 연구에서는 미리 학습시킨 심층 컨볼루션 신경망에 계층을 추가하여 YOLO (You Only Look Once) 네트워크를 설계하였다. 이를 통해 객체 탐지 문제에 있어 기존의 회귀 기반 탐지 알고리즘이 아닌 사각형 영역 라벨링을 통해 탐지된 물체의 클래스와 탐지 정밀도를 출력하는 시스템을 구현하였다. [3]의 YOLO 모델은 초당 45프레임의 실시간 이미지 처리도 가능하며, 고정밀 객체탐지가 가능하다.

[4]의 논문에서는 미리 학습된 네트워크를 활용하는 전이학습 기법을 이용하여 Human Epithelial-2 이미지의 특징 및 패턴을 학습하여 분류하는 알고리즘을 제안하였다. 제안한 시스템은 입력 이미지 데이터의 특징을 추출하여 1,000개의 클래스로 분류하는 기능을 수행한다. 제안한 시스템을 통해 최고 정밀도 91.5%를 얻었고, ICPR2012 데이터셋에 대해서는 전문가가 분류한 정밀도 보다 약 4% 높은 정밀도 성능을 기록하였다.

[5]의 연구에서는 전이학습을 활용한 Gaussian 프로세스 회귀 기반 객체 추적 알고리즘을 제안하였다. 해당 연구에서는 기존의 객체 분류 알고리즘에서 분류 정밀도를 확률값으로 출력하는 형태가 아닌 Gaussian 프로세스 회귀기법을 활용하여 이 확률을 분석하여 성능개선을 이루었다.

인공지능 기반 영상인식 시스템에서 높은 정밀도를 갖는 객체탐지 네트워크를 학습시키기 위해 많은 양의 데이터가 필요하게 된다. 본 논문에서는 [3]에서 제안한 YOLO 네트워크를 활용하여 [3]의 연구에서 사용한 학습데이터의 약 20% 만을 이용하여 높은 정밀도를 나타내는 전이학습 기반 객체탐지 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 전이학습의 기본 개념과 전이학습 네트워크에 대해 기술한다. 3장에서는 실험에 사용한 데이터와 설정된 실험 파라미터에 대해 설명하고 이에 따른 실험결과와 성능지표를 제시한다. 마지막 4장에서는 논문을 마무리하며 결론을 제시한다.

II. 딥 러닝 알고리즘

1. 전이 학습

전이학습은 기계학습 분야에서 부족한 학습데이터의 문제를 보완할 수 있는 수단으로써, source domain에 대한 지식을 target domain으로 전이시켜 적은 학습데이터 상황에서 모델을 학습시키는 기법이다^[1, 2]. 그림 1

을 보면 전이학습의 기본구조를 나타내고 있다. 그림을 보면 먼저 source domain에서 데이터를 통해 네트워크를 사전 학습시키게 된다. 이후 전이학습을 적용하기 위해 source domain 네트워크에 추가적으로 계층을 추가하여 transfer learning 네트워크를 구성한다. 이후 target domain 데이터를 활용해 전이학습 네트워크를 학습시켜 새로운 target domain에 대한 출력을 얻게 된다. 이 일련의 과정을 전이학습 과정이라고 하며, 사전 학습된 source domain 네트워크에 새로운 계층을 추가하여 target domain에서 전이학습 모델을 학습시키는 것을 fine-tuning 이라고 한다^[6].

전이학습은 이처럼 사전 학습된 네트워크의 지식이 새로운 target domain으로 전이되어 적은 데이터로도 좋은 성능을 나타내는 특징이 있다.

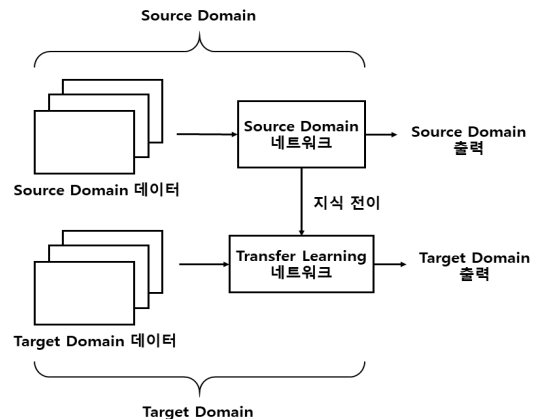


그림 1. 전이학습 기본구조.
Fig. 1. Basic structure of transfer learning scheme.

2. 전이학습 네트워크

본 논문에서는 YOLO 네트워크와 Resnet-50을 활용한 전이학습 네트워크를 사용하였다. 먼저 YOLO 네트워크는 24개의 컨볼루션 계층과 2개의 퍼셉트론 계층으로 이루어진 네트워크이며, ImageNet 데이터셋을 사전 학습시킨 네트워크이다^[3]. YOLO 네트워크는 출력으로써 탐지한 객체에 대해 복수의 사각형 경계값을 추정하게 된다.

다음으로 Resnet-50은 50개의 계층으로 이루어진 컨볼루션 신경망이며, YOLO 네트워크와 마찬가지로 ImageNet 데이터셋으로 사전 훈련이 진행된 네트워크를 활용하였다^[7]. Resnet-50의 출력은 1,000개의 클래스로 나누어진 분류출력을 나타내게 된다.

본 논문에서는 Resnet-50을 기본 네트워크로 설정하고 Resnet-50의 특징 추출 계층에 YOLO 네트워크를 연결하여 전이학습 네트워크를 구성하였다.

III. 실험 환경 및 결과

1. 실험 데이터

본 논문에서는 Leeds Sports Pose 데이터셋의 스포츠 동작을 취하고 있는 사람의 이미지 데이터를 이용하였다^[8]. 이 데이터셋은 12,000 장의 이미지로 구성되어 있다. 본 논문에서는 12,000 장의 데이터 중 2,000 장의 데이터를 사용한다. 학습데이터로 1,9000 장의 이미지를 이용하였고, 테스트 데이터로 100 장의 이미지를 사용하여 실험을 진행하였다.

모든 데이터의 입력 크기는 224x224 픽셀로 재조정하였다. 모든 이미지 데이터에는 1명 이상의 사람이 나타나 있으며, 이미지에 나타난 사람의 수가 2명 이상일 경우 더 가까운 곳에 있거나, 더 많은 넓이를 포함하고 있는 사람을 탐지하도록 학습을 진행하였다.

2. 실험 환경

본 연구에서 설정한 실험의 파라미터는 mini-batch 사이즈, 학습률, epoch 총 3가지다. 각 파라미터에 대해 간략히 서술하겠다.

먼저 mini-batch 사이즈는 인공지능 모델이 학습을 진행할 때, 한 차례에 입력받아 연산할 데이터의 단위로, 이 값은 16으로 설정하였다. Epoch는 학습 데이터에 대한 반복학습 횟수이다. 본 연구에서는 이 파라미터를 20으로 설정했다.

마지막으로 본 논문에서는 optimizer로써 기본적인 stochastic gradient descent method를 사용하였다^[9]. 최적화 기법에 따른 학습률은 0.001로 설정하였다.

3. 실험 결과

그림 2는 전이학습을 통해 학습시킨 YOLO 네트워크 기반 객체 탐지 알고리즘의 객체탐지 결과를 보여준다. 사각형 영역은 객체 탐지 네트워크가 사람이 있다고 판단한 영역을 나타내고, 영역 위의 숫자는 네트워크가 사람이 있는 영역을 판별할 때 갖는 신뢰도를 나타낸다. 그림을 보면 두 결과 모두 2명 이상의 사람이 있지만 더 카메라에 가까운 사람을 탐지하는 결과를 확인할 수 있다.

왼쪽 그림에 비해 오른쪽 그림은 2명의 사람이 한 영역에 더 많은 넓이의 픽셀을 공유하고 있기 때문에, 네트워크의 score가 왼쪽에 비해 더 낮게 책정된 것을 볼 수 있다.

그림 3은 학습된 객체 탐지기에 대한 recall 대비 precision 그래프를 보여준다. Recall은 탐지기가 모든 샘플에 대해 탐지를 실패하지 않고 탐지했는지를 나타내는 파라미터이고, precision은 탐지한 결과가 얼마나 정확한지 나타내는 파라미터이다^[10]. 그림 3을 보면 recall의 경우 84%로, 2명 이상의 사람이 있는 경우와 사람의 형태가 모호한 경우 탐지 성능이 떨어지는 것을 확인하였지만, precision값은 97.67%로 높은 수준을 기록하였다.



그림 2. 객체 탐지결과 예시.
 Fig. 2. Object detection examples.

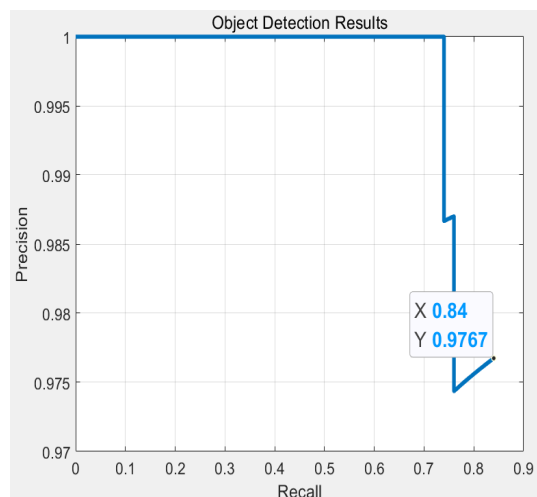


그림 3. 객체 탐지결과.
 Fig. 3. Object detection results.

IV. 결 론

본 논문에서는 YOLO 네트워크를 활용한 전이학습을 적용하여 적은양의 데이터를 통해 새로운 객체 탐지 네트워크를 학습시켰고, 이는 YOLO 네트워크를 학습시킨 데이터의 20% 수준이다. 또한, 이를 통해 97.67%의 정밀도를 갖는 사람 신체 탐지기를 도출하였다. 추후 연구로는 recall의 값을 높이기 위해 복수의 객체를 탐지하는 네트워크를 설계하는 것을 고려할 수 있다.

References

- [1] I. Goodfellow, *Deep Learning*, MIT Press, 2016.
- [2] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang and C. Liu, "A survey on deep transfer learning," *Proc. The 27th International Conference on Artificial Neural Networks*, pp. 270-279, Rhodes, Greece, Oct. 2018.
- [3] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE International Conference Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, Las Vegas, NV, USA, June 2016. DOI: 10.1109/CVPR.2016.91
- [4] H. T. H. Phan, A. Kumar, J. Kim and D. Feng, "Transfer learning of a convolutional neural network for HEP-2 cell image classification," *Proc. 2016 IEEE 13th International Symposium on Biomedical Imaging*, pp. 1208-1211, Prague, Czech, Apr. 2012. DOI: 10.1109/ISBI.2016.7493483
- [5] J. Gao, H. Ling, W. Hu and J. Xing, "Transfer learning based visual tracking with Gaussian process regression," *Proc. of ECCV 2014*, pp. 188-203, Zurich, Switzerland, Sept. 2014. DOI: https://doi.org/10.1007/978-3-319-10578-9_13
- [6] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, "How transferable are features in deep neural networks?," *Proc. Advances in Neural Information Processing Systems*, pp.3320-3328, Montreal, Canada, Dec. 2014.
- [7] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, Las Vegas, NV, June 2016. DOI: 10.1109/CVPR.2016.90
- [8] S. Johnson and M. Everingham, "Learning effective human pose estimation from inaccurate annotation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR2011)*, pp. 1465-1472, Providence, RI, USA, June 2011. DOI: 10.1109/CVPR.2011.5995318
- [9] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv:1609.04747, June 2017.
- [10] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," *Proc. International Conference Machine Learning*, pp. 233-240, Pennsylvania, USA, June 2006.
- [11] J. Y. Kim and P. G. Lee, "Image recognition based on adaptive deep learning," *Journal of the Institute of Internet Broadcasting and Communication*, vol.18, no.1, pp.113-117, Feb. 2018.

저 자 소 개

이 동 구(준회원)



- 2019년 2월 : 광운대학교 전자융합공학 학사 졸업
- 2019년 3월~현재 : 광운대학교 전자융합공학과 석박사통합과정
- 관심분야 : 인공지능, 위성통신, 스마트 그리드

선 영 규(준회원)



- 2018년 2월 : 광운대학교 전자융합공학 학사 졸업
- 2018년 3월~현재 : 광운대학교 전파공학과 석박사통합과정
- 관심분야 : 인공지능, 무선 에너지하베스팅, 디지털통신, 전력선 통신

김 수 현(준회원)



- 2019년 2월 : 광운대학교 전자융합공학 학사 졸업
- 2019년 3월~현재 : 광운대학교 전자융합공학과 석박사통합과정
- 관심분야 : 인공지능, 전력선통신, 스마트 그리드

심 이 삭(준회원)



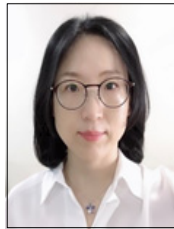
- 2016년 2월 : 광운대학교 전자융합공학과 졸업
- 2016년 3월~ 현재 : 광운대학교 전파공학과 석박통합과정
- 관심분야 : 전력선 통신, 무선 에너지 하비스팅, Backscatter, 인공지능

이 계 산(정회원)



- 2002년 2월: 게이오대학교 전자공학과 공학박사
- 2002년~2003년 : 일본 KDDI 연구원
- 2003년~ 현재: 경희대학교 전파공학과 교수
- 관심분야 : DS-CDMA, OFDM, LED-ID, LTE-A

송 명 남(준회원)



- 2001년 2월 : 서울여자대학교 컴퓨터학과 석사
- 2001년 ~ 2012년 : (주)XTM 월드 전자팀 차장
- 2012년 ~ 현재 : (주)현대엠아이비인 터내셔널 기술개발팀 팀장

- 관심분야 : 영상처리, 신호처리, 임베디드 시스템

김 진 영(정회원)



- 1998년 2월 : 서울대학교 전자공학과 공학박사
- 2001년 2월 : SK텔레콤 네트워크 연구소 책임연구원
- 2001년 3월~현재 : 광운대학교 전자융합공학과 교수

- 관심분야 : 차세대이동통신, 가시광통신, 전력선통신, 인공지능

※ 위 논문은 문화체육관광부의 스포츠산업기술개발사업에 의해 국민체육진흥공단의 국민체육진흥기금을 지원받아 연구되었습니다.