

비디오에서 YOLOv3 기반 차량 인식 및 계수 방안

이혜진[†], 이은지[†], 박소현[†], 임선영^{††}, 박영호^{†,*}

[†]숙명여자대학교 IT 공학과

^{††}숙명여자대학교 빅데이터 활용 연구센터

e-mail : {adorablehye96, iej9031, shpark, sunnyihm, yhpark}@sookmyung.ac.kr

*교신저자

YOLOv3-based Vehicle Detection and Counting Method through Video

Hye-Jin Lee[†], Eun-Ji Lee[†], So-Hyun Park[†], Sun-Young Ihm^{††}, Young-Ho Park^{†,*}

[†]Dept. of IT Engineering, Sookmyung Women's University

^{††}Bigdata Using Research Center, Sookmyung Women's University

요 약

본 논문에서는 YOLOv3 기반의 차량 인식 및 계수 방안 연구를 진행한다. 이를 위해, 단일 단계 방식의 구조를 가진 YOLOv3 를 활용하여 영상 속 차량 인식 및 계수 하는 실험을 수행한다. 위의 실험을 통해 영상 속 차량의 수와 밝기에 인식을 정도와 차량 계수 평균 값 추출을 하였다. 실험 결과, 영상 속의 차량의 수와 밝기에 상관없이 90%이상의 차량에 대한 높은 인식률을 보였으며 1 초 당 약 20 번씩 인식되는 차량의 수에 대한 계수의 평균 값을 추출하여 텍스트 파일 형태로 저장하여 시간에 따른 차량 계수를 확인하였다. 본 기술은 현대 사회의 교통 체증을 해결하기 위한 교통 흐름 예측에 활용될 수 있다.

1. 서론

최근 IT 기술의 빠른 발달과 함께 기존 교통 시스템에 첨단 기술을 접목시키는 ITS(Intelligent Transport System) 기술을 활용하여 차량 및 교통량으로 인한 교통 정체, 사고 발생 등의 각종 교통 문제를 체계적으로 대처하여 효율적인 교통 운영과 교통 이용 편의를 높이고자하는 다양한 연구들이 등장하고 있다. 다양한 분야 중 영상을 통한 차량 인식 및 계수를 통한 교통 흐름 파악 연구를 진행한다.

오늘날의 CCTV 의 역할은 단순 감시가 아닌 범죄 예방, 매장을 방문하는 고객의 수, 고객의 동선 파악 등 마케팅에 다양하게 활용되는 사례가 늘어나며 CCTV 의 역할은 나날이 커져가고 있다. 다양한 CCTV 의 역할 중 계수하는 역할을 차량에 접목시켜 영상 속 차량의 수를 파악하여 교통 흐름 예측을 한다. 이를 통해, 현대 사회의 교통 정체를 해결하여 효율적인 교통 운영과 교통 이용 편의를 높이고자 한다.

따라서, 본 논문에서는 CCTV 영상 속 차량 인식 및 계수를 통한 교통 흐름 분석 및 예측하는 방안을 제안하고자 한다. 이 방법은 YOLOv3 기반으로 차량 인식 및 계수를 처리하는 방법이다. YOLOv3 알고리즘은 2.2 절에서 설명한다.

본 논문의 구성은 다음과 같다. 2 장에서는 대표적인 객체 인식 알고리즘들에 대해 설명한다. 3 장에서

는 제안하는 방법을 설명한다. 4 장에서는 실험에 대하여 설명하는데, 먼저 개발 환경을 소개하고 후에 실험 및 결과 분석을 진행한다. 마지막으로 5 장에서는 결론 및 향후 연구에 대해 소개한다.

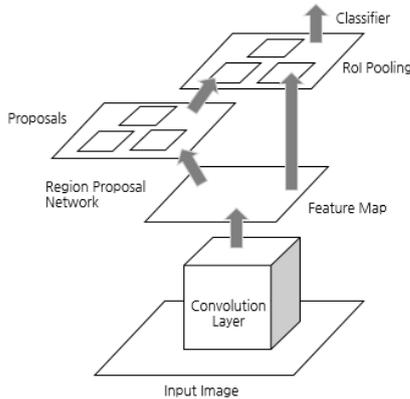
2. 관련 연구

본 장에서는 대표적인 객체 인식 알고리즘에 대하여 크게 2 가지로 나누어 설명한다. 2.1 절에서는 2 단계 방식 구조의 객체 인식 알고리즘인 Faster R-CNN 알고리즘에 대하여 설명하고, 2.2 절에서는 단일 단계 방식 구조의 객체 인식 알고리즘인 SSD(Single shot Detector) 알고리즘에 대하여 설명한다.

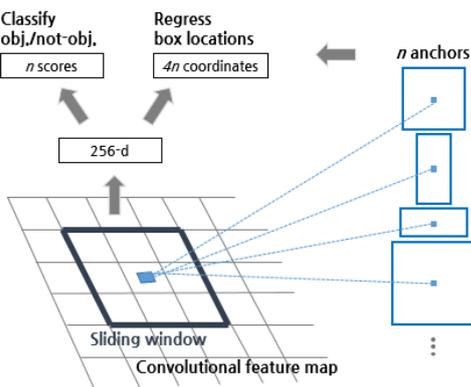
2.1 Faster R-CNN 의 아키텍처

Faster R-CNN 은 네트워크에서 생성한 피쳐 맵 위에 $n * n$ 크기의 윈도우를 생성하여 슬라이딩하며 Region proposals 을 생성하는 FCN(Fully Convolutional Network)형태의 RPN(Region Proposal Network)을 도입하여 기존의 Fast R-CNN 을 개선한 알고리즘이며[1, 5], 상대적으로 속도가 느린 CPU 로 계산하는 대신 빠른 RPN(GPU)을 사용하여 성능을 개선하였다. Faster R-CNN 의 아키텍처는 그림 1 을 통해 확인 가능하다. RPN 는 이미지를 입력 받아 사각형 형태의 오브젝트 프로포잘과 객관적인 점수를 출력해주는 역할을 한다.

그림 2[6]를 보면 동일한 크기의 슬라이딩 윈도우를 이동시키며 윈도우의 위치를 중심으로 사전에 정의된 다양한 비율과 크기로 이루어진 앵커 박스들을 적용하여 특징을 추출한다[6]. 그러므로 피라미드 특징 계층 구조처럼 이미지 크기를 조절할 필요가 없으며 여러 규모의 슬라이딩 윈도우처럼 필터 크기를 변경할 필요도 없으므로 계산 효율이 높다는 장점을 갖고 있다. 하지만, 2 단계 방식 구조는 공통적으로 복잡한 파이프라인을 가지며 속도가 느려 실시간 객체 감지가 힘들다는 단점을 갖고 있다.



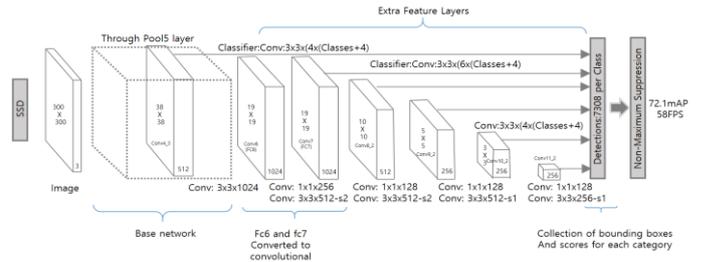
(그림 1) Faster R-CNN의 아키텍처



(그림 2) RPN의 아키텍처

2.2 SSD(Single Shot Detector)의 아키텍처

SSD는 VGG16[4] 네트워크를 기능 추출기로 사용하는 단일 샷 검출기이다. 이 알고리즘은 그림 3[2]과 같이 피쳐 맵을 뽑아내는 과정까지 하나의 네트워크 구조를 가지는 알고리즘이다. SSD는 피쳐 맵을 여러 개의 크기로 만들어서, 큰 피쳐 맵에서는 작은 물체의 검출을 하고 작은 피쳐 맵에서는 큰 물체의 검출을 하여 위치 추정 및 입력 이미지의 재표집을 없애면서 정확도 높은 결과를 도출한다. 여러 피쳐 맵은 한 이미지를 다양한 그리드로 접근하여 다양한 크기의 물체들을 검출할 수 있도록 하며[2], 큰 물체는 비율이 달라져도 잘 찾는 장점을 갖고 있다. 하지만, 여러 개의 피쳐 맵의 검출을 다 계산하므로 계산 비용이 증가할 수 있으며 입력 사이즈가 크지 않다면 최상위 계층에서부터 정보가 적기 때문에 작은 물체는 잘 찾지 못한다는 단점을 갖고 있다.



(그림 3) SSD의 아키텍처

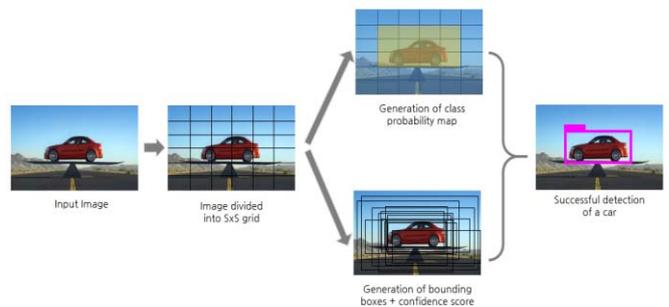
3. 제안하는 CCTV 차량 계수 방법

본 장에서는 앞서 2장에서 설명했던 관련 연구들보다 실시간 객체 인식 알고리즘 중에서 본 연구에 제일 적합하다고 판단되는 YOLOv3 알고리즘을 활용하여 CCTV에서 차량 인식 및 계수 방안에 관하여 제안하는 방법을 설명한다. 먼저, 3.1 절에서는 YOLOv3 알고리즘에 대하여 설명하고, 3.2 절에서는 YOLOv3 알고리즘을 이용한 차량 인식 및 계수 방안에 대하여 설명한다.

3.1 YOLOv3의 아키텍처

YOLOv3는 그림 4[8]와 같이 단일 신경망을 적용하여 이미지를 일정한 영역으로 분할하고 각각의 영역에 가중치를 부여하는 방식을 사용한다. 객체를 여러 번 호출하는 기존의 객체 인식 알고리즘과는 다르게 이미지를 한 번만 호출하므로 속도 측면에서 R-CNN보다 1000 배 이상 빠르며 Fast R-CNN보다 100 배 빠른 속도와 높은 정확도를 보이는 알고리즘이다.

FPN(Feature Pyramid Network)의 컨셉과 유사한 방식으로 각각 2 배씩 차이가 나는 3 개의 다른 스케일의 박스들을 예측하고 다양한 스케일에서 특징들을 추출한다. 3 개의 바운딩 박스 * 3 개의 피쳐 맵으로 이루어진 총 9 개의 앵커 박스는 k-means 클러스터링을 통해 결정된다[3]. 다른 알고리즘들의 비해 간단한 처리 과정으로 속도가 매우 빠르며 이미지 전체를 한 번에 바라보는 방식으로 클래스에 대한 맥락적 이해도가 높다는 장점을 갖고 있다. 하지만, 작은 오브젝트에 대하여 상대적으로 낮은 정확도를 보이는 단점을 갖고 있다.



(그림 4) YOLOv3의 아키텍처

3.2 YOLOv3 기반의 차량 인식 및 계수 방안

본 절에서는 비디오에서 YOLOv3 기반 차량 인식 방안을 제안한다. 제안하는 방법은 YOLOv3 기반에서 차량을 인식 및 계수하는 단계에 대하여 설명한다. 단계는 총 세 단계로 구성된다.

- a) Step 1 (단일 신경망을 통한 영역 분할 단계): 단일 신경망을 통한 영역 분할 단계에서는 입력 받은 이미지를 단일 신경망을 적용하여 일정하게 분할한다.
- b) Step 2 (경계 상자 및 확률 예측 단계): 경계 상자 및 확률 예측 단계에서는 로지스틱 회귀를 이용하여 각 경계 상자에 대한 객관적인 점수를 예측하고 앵커 박스와 실제 바운딩 박스의 제일 좋은 IoU 에 값을 1 로 두어 계산하여 경계 상자 및 예측을 한다.
- c) Step 3 (데이터 추출 및 저장 단계): 데이터 추출 및 저장 단계에서는 이미지에서 추출한 객체에 경계 상자 및 확률 예측이 표시되어 저장된다. 동시에 현재 시간과 함께 영상 속에서 인식된 차량의 수가 텍스트 파일로 저장된다.

이때, 영상 속에서 인식된 차량은 1 초에 약 20 번 씩 텍스트 파일로 저장된다. 1 초 당 인식된 차량의 수에 대한 오차 범위를 줄이기 위해 1 초 당 인식된 차량의 수에 대한 평균 값을 추출하여 새로운 텍스트 파일로 다시 저장하였다. 이를 통해, 시간에 따른 차량 계수 결과를 확인하였으며, 결과 이미지는 4.2 절 실험 및 결과 분석에서 확인할 수 있다.

4. 실험

본 장에서는 개발 환경과 실험 결과에 대해 설명한다. 먼저, 4.1 절에서는 개발 환경을 설명하고, 4.2 절에서는 실험 및 결과 분석을 설명한다.

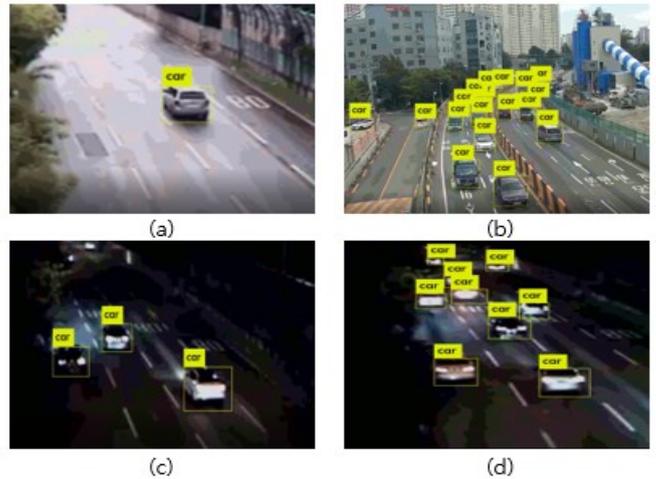
4.1 개발 환경

본 연구에서는 경찰청 도시교통정보센터에서 제공하는 CCTV[7]를 사용하였다. 실험 영상은 차량의 수가 적고 많을 때와 영상이 밝고 어두울 때에 따른 인식률을 비교하기 위해 출퇴근 시간 때인 오전 7 시부터 오전 9 시와 오후 6 시부터 오후 8 시까지 2 가지 상황으로 나누어 진행하였다. 개발을 진행한 컴퓨터 사양은 Intel® Core™ i7-2000 CPU @ 3.40GHz 이고 그래픽 카드는 GeForce GTX 1080 Ti 를 사용하였다.

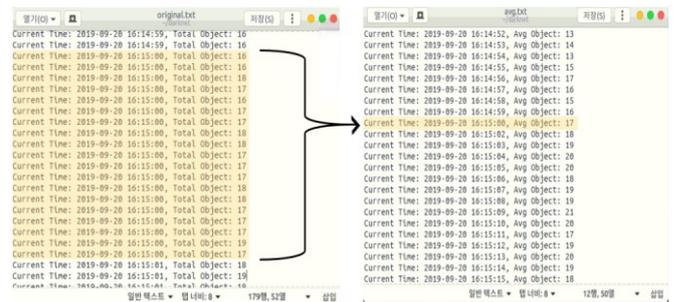
4.2 실험 및 결과분석

그림 5 는 차량의 수가 적고 많을 때와 영상이 밝고 어두울 때에 따른 인식률을 비교하기 위해 출퇴근 시간 때인 오전 7 시부터 오전 9 시와 오후 6 시부터 오후 9 시로 나누어 실험을 진행하였다. 그림 5 의 (a), (b)는 출근 시간 때인 8 시 30 분쯤의 이미지이고 그림 5 의 (c), (d)는 퇴근 시간 때인 오후 8 시쯤의 이미지이다. 그림 5 의 (a), (c)는 차량의 수가 적은 경우이고 그림 5 의 (b), (d)는 차량의 수가 많은 경우이다.

실험 결과, 차량 인식은 영상 속 차량의 수, 영상의 밝기와 상관없이 각 차량에 대한 검증 정확도가 평균 90%이상으로 높은 인식률을 보였다. 그 결과를 1 초 당 약 20 번씩 인식되는 차량의 수에 대하여 오차 범위를 줄이기 위해 1 초당 인식된 차량의 수에 대한 평균 값을 추출하여 새로운 텍스트 파일로 저장하여 시간에 따른 차량 계수를 확인하였다. 그림 6 의 좌측은 원본 텍스트 파일이며 우측은 인식된 차량의 수에 대한 평균 값을 추출하여 저장한 텍스트 파일의 결과 화면이다.



(그림 5) 영상 속 차량 인식



(그림 6) 차량 수의 평균 값 텍스트 파일

5. 결론 및 향후 연구

본 논문에서는 CCTV 의 다양한 역할 중 물체를 인식하여 계수하는 역할을 차량과 접촉시켜 영상 속 차량을 인식 및 계수하여 교통 흐름 예측을 통해 현대 사회의 교통 체증을 해결하기위해 교통 흐름 예측을 하는 방안을 제시하였다. 데이터는 도시교통정보센터에서 제공하는 CCTV 영상을 수집하여 사용하였으며 객체 인식 알고리즘은 YOLOv3 를 적용하였다.

실험 결과, 영상 속 차량의 수 및 영상의 밝기와 상관없이 차량에 대한 검증 정확도가 평균 90%이상으로 좋은 인식률을 보였다. 이를 통해, 1 초당 약 20 번씩 인식된 차량의 수에 대한 오차 범위를 줄이기 위해 1 초당 인식된 차량의 수에 대한 평균 값을 추출한 결과를 텍스트 파일로 저장하여 교통 흐름 예측을 하고자 하였다. 하지만 영상의 모서리 쪽으로 갈

수록 인식이 잘 안된다는 문제점과 차량의 수가 급격하게 많아지면 같은 색의 차량이 하나의 차량으로 인식되는 문제점에 대하여 보완 방법이 필요하다.

향후 연구로는 영상 모서리 쪽의 인식을 개선과 차량의 수가 많아질 경우 같은 색의 차량이 하나의 차량으로 인식되는 문제 개선 연구 등을 진행할 예정이다.

사시문구

이 논문은 2019 년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2016-0-00406. (기반 SW-창조씨앗 2 단계)SIAT 형 CCTV 클라우드 플랫폼 기술 개발) 그리고 2019 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2018-0-00225, 과학적 정책 수립을 위한 도시행정 디지털트윈 핵심 기술 개발)

참고문헌

- [1] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar and K.Ouni, "Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3," in Proc. 1st Int. Conf. Unmanned Vehicle Syst.-Oman (UVS), 2019.
- [2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed, "SSD:Single shot multibox detector," arXiv:1512.02325, 2015.
- [3] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
- [4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations (ICRL), 2015.
- [5] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision, 2015.
- [6] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," In NIPS, 2015.
- [7] CCTV 영상 제공,
<https://www.utic.go.kr:449/main/main.do>
- [8] 자동차 사진, <https://news.join.com/article/21793292>