

딥러닝을 이용한 보행자 검출 알고리즘

곽내정*, 송특섭**

*충북대학교 정보통신공학과

**목원대학교 컴퓨터공학과

e-mail : knj0125@chungbuk.ac.kr

Pedestrian detection algorithm using Deep Neural Network

Nae-Joung Kwak*, Teuk-Seob Song**

*Dept of Information & Comm. Engineering, Chungbuk National University

**Dept of Computer Science, Mokwon University

요 약

본 연구는 ResNet에 국부적 특성의 검출 성능이 뛰어난 Grad-CAM을 결합하여 보행자를 검출하는 알고리즘을 제안한다. 또한 폐색 영역(가려진 영역)이 있는 보행자의 검출을 위해 폐색 영역이 존재하는 데이터로 훈련하여 폐색 영역이 존재하는 경우에도 검출 성능이 강건한 모델을 제안한다.

1. 서론

보행자 검출은 영상 내 서 있거나 걷고 있는 사람의 영역을 검출하는 기술로 수년간 가장 광범위하게 연구된 문제 중 하나이다. 보행자 검출은 보행자 들이 각자 다른 모습과 자세를 갖고 있으며 물체나 다른 보행자에 의해 가려진 경우에도 검출할 수 있어야 한다. 보행자 검출 기법은 Hand-crafted model과 Deep model로 나누어 지며 최근 보행자 검출은 Deep model을 활용한 연구로 발전하고 있다. 다양한 Deep model 중 ResNet[1]은 residual network과 shot-cut connection을 이용하여 152개의 layer를 사용하여 빠른 속도로 영상을 분류하는 방법을 제안했다. 또한 Faster R-CNN 기법[2]을 적용하여 영상의 지역 성과 검출 성능을 개선했다.

본 연구는 ResNet을 기반으로 영상의 국부적 특징을 추출하는 방법을 적용하여 보행자 검출에 적용하고자 한다. 영상의 국부적 특징 추출을 위해 Grad-CAM[3]을 결합하여 특징을 분석한다. 또한 폐색 영역(가려진 영역)이 있는 보행자의 검출을 위해 폐색 영역이 존재하는 데이터로 훈련하여 폐색 영역이 존재하는 경우에도 강건한 모델을 제안한다.

2. 기존의 방법

2.1 ResNet

ResNet은 네트워크의 깊은 구조에 비례하는 학습성능을 높이기 위해 Residual Learning이라는 학습 방법을 제안

했다. 이것은 CNN의 구조를 변형하여 입력에서 바로 출력으로 연결하는 연결을 생성하며 그림1에서 $F(x)$ 가 0이 되는 방향으로 학습하면 입력의 작은 움직임 검출할 수 있으며 입력의 작은 움직임 즉, 나머지(residual)을 학습한다. 또한 입력의 x 가 바로 출력으로 연결되므로 파라미터의 수에 영향이 없고 한 번의 덧셈 외에 연산량의 증가는 없다.

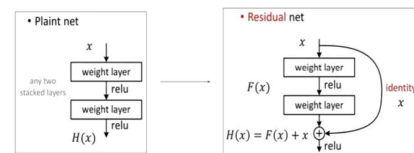


그림 1. Residual Network의 구조

또한 몇 개의 레이어를 건너뛰며 입력과 출력이 연결되므로 forward 및 backward 경로가 단순해진다. 이러한 이유로 깊은 망도 단순해지며 늘어난 길이로 인해 좀 더 정확한 성능을 얻을 수 있다.

객체의 검출은 그림2와 같이 특징 추출과 객체 검출로 나눌 수 있는데[4] 객체 분류 뿐은 CNN을 통해 얻어진 특징을 기반으로 별도의 망으로 처리한다는 개념으로 NoC(network on Conv feature map)이라하였다.

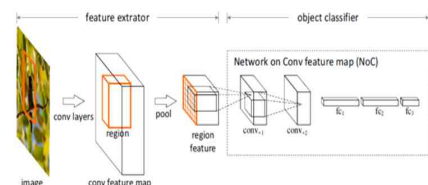


그림 2. NoC의 개요

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.NRF-2018R1D1A1B07048675)

정확도를 개선하기 위해 FC 레이어(fully connected layer) 앞에 컨볼루션 레이어를 추가하고 multi-scale을 대응하기 위해 인접한 scale에서의 특징맵 결과를 선택하는 Maxout 방식을 도입한다.

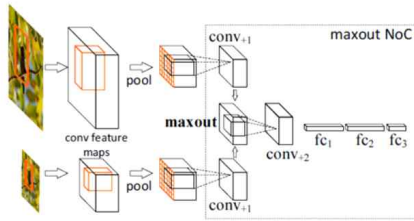


그림 3. maxout NoC의 예

2.2 Grad-CAM

Grad-CAM은 CAM[5]의 단점을 개선한 모델이다. CAM은 FC 레이어를 GAP(Global Averager Pooling)로 교체하고 fine-tuning함으로써, 뉴럴 네트워크가 이미지의 어떤 부분을 보고 특정 레이블로 판단을 내리는지에 대해 알 수 있도록한다. 그러나 FC를 GAP로 대체해야 한다는 점, GAP 직전의 컨볼루션 레이어에서만 쓸 수 있다는 점, 또 GAP 뒷단에 있는 Dense Layer의 가중치 정보가 필요하므로 fine-tuning이나 re-training의 과정을 거쳐야하는 등의 단점이 있다. 이것을 개선하여 CAM을 일반화하여 객체 검출 외에 Visual Question Answer(VQA)나 Captioning처럼 다양한 목적을 수행하는 CNN에 CAM을 적용할 수 있도록하였으며 FC를 사용한 기존 모델 아키텍처와 마지막 conv layer가 아닌 다른 레이어에서도 Grad-CAM을 적용할 수 있도록 하였다. Grad-CAM은 영상의 국부적 특성을 강조하고 시각화하는 성능이 있다.

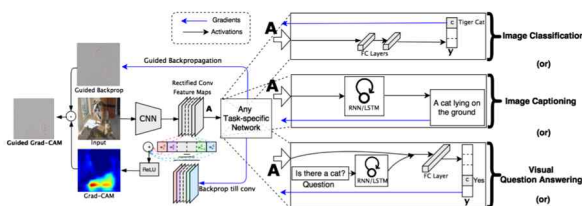


그림 4. Grad-CAM의 개요

3. 제안방법

본 연구는 ResNet이 컨볼루션 레이어 및 NoC를 이용하여 Faster R-CNN의 개념을 적용한 것을 응용하여 ResNet에 Grad-CAM을 접목하여 개선된 객체 검출 성능을 보이는 네트워크를 구현하고자 한다.

기존의 ResNet에서 생성되는 특징맵의 후반부의 일정 영역에 Grad-CAM을 적용하여 ResNet의 레이어에서 보행자의 국부적 특성 변화를 고찰하고 NoC와의 성능을 비교한다.

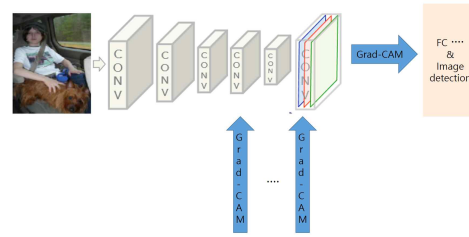


그림 5. ResNet에 Grad-CAM의 적용 예

또한 폐색 영역의 검출 성능을 높이기 위해 폐색 영역 데이터를 학습데이터로 사용하여 폐색영역에도 강건한 네트워크를 구성하도록 한다. 이때 수동적으로 학습 데이터를 만들 경우 비용이 많이 들고 폐색의 수준에 제한이 있을 수 있다. 또한 어느 시점에는 폐색 영역을 설정하는 박스나 크기의 위치가 동일하게 될 수 있다. 본 연구에서는 [6]의 알고리즘을 적용하여 폐색영역을 설정한다.

4. 결론

본 논문에서는 ResNet을 기반으로 하여 보행자를 추출하는 알고리즘을 제안하고자 한다. 제안 알고리즘은 차후 실제 구현하고 보행자 데이터셋을 통해 제안 알고리즘의 성능을 검증하고자 한다.

참고문헌

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, pp. 91-99, 2015.
- [3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *IEEE International Conference on Computer Vision*, pp. 618-626, 2017.
- [4] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "object detection networks on convolutional feature maps," *Journal of IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 39, pp. 1476-1481, 2016.
- [5] B. Zhou, A. Khosla, L. A., A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921-2929, 2016.
- [6] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random Erasing Data Augmentation," *Arxiv preprint arXiv:1708.04896v2*, 2017.