

딥러닝 기반의 회전에 강인한 텍스트 검출 기법

*최인규 **김제우 ***송혁 ***유지상

전자부품연구원

*cig2982@keti.re.kr

Rotation-robust text localization technique using deep learning

*Choi, In-Kyu **Kim, Jewoo **Song, Hyok ***Yoo, Jisang

*, **Korea Electronics Technology Institute, ***Kwangwoon University

요약

본 논문에서는 자연스러운 장면 영상에서 임의의 방향성을 가진 텍스트를 검출하기 위한 기법을 제안한다. 텍스트 검출을 위한 기본적인 프레임 워크는 Faster R-CNN[1]을 기반으로 한다. 먼저 RPN(Region Proposal Network)을 통해 다른 방향성을 가진 텍스트를 포함하는 bounding box를 생성한다. 이어서 RPN에서 생성한 각각의 bounding box에 대해 세 가지의 서로 다른 크기로 pooling된 특징지도를 추출하고 병합한다. 병합한 특징지도에서 텍스트와 텍스트가 아닌 대상에 대한 score, 정렬된 bounding box 좌표, 기울어진 bounding box 좌표를 모두 예측한다. 마지막으로 NMS(Non-Maximum Suppression)을 이용하여 검출 결과를 획득한다. COCO Text 2017 dataset[2]을 이용하여 학습 및 테스트를 진행하였으며 주관적으로 평가한 결과 기울어진 텍스트에 적합하게 회전된 영역을 얻을 수 있음을 확인하였다.

1. 서론

길거리의 이정표, 가게의 이름, 물건의 이름과 같은 실생활의 장면 영상 내에 있는 텍스트는 환경에 대한 필수적인 정보를 포함하고 있기 때문에 중요한 역할을 할 수 있다. 텍스트 기반의 검색, 번역 등 다양한 분야에 활용하기 위해서는 텍스트 검출과 인식이라는 핵심적인 두 가지 과정이 필요하다. 본 논문에서는 영상 내 텍스트의 위치를 예측하는 텍스트 검출에만 초점을 맞춘다. 장면 텍스트 검출은 텍스트의 크기, 너비와 높이 비율, 글꼴, 조명, 왜곡, 방향이 다르기 때문에 어려움이 있다. 일반적인 객체 검출과는 다르게 텍스트의 방향 정보가 텍스트 인식이나 다른 분야 적용에 유용하기 때문에 정렬된 bounding box 예측뿐만 아니라 텍스트의 방향도 예측할 필요가 있다.

본 논문에서는 텍스트 방향에 맞춰 회전된 bounding box 획득을 목표로 한다. 텍스트 검출을 위한 기본적인 프레임워크는 Faster R-CNN을 이용하고 특징 추출을 위한 CNN은 ResNet-101[3] 구조를 이용한다. 먼저 RPN(Region Proposal Network)을 통해 텍스트를 포함하는 bounding box를 예측한다. 그리고 각각의 예측된 bounding box에 대해서 세 가지의 서로 다른 크기의 pooling된 특징지도를 추출한다. 이는 일반적인 객체 검출과는 다른 텍스트 검출 문제를 위한 개선이며 수평 및 수직 방향의 텍스트 검출 성능을 향상시킨다. Pooling된 특징지도들을 병합하고 텍스트와 텍스트가 아닌 대상에 대한 score, 정렬된 bounding box 좌표, 회전된 bounding box 좌표를 동시에 예측한다. 마지막으로 NMS(Non-Maximum Suppression)을 이용하여 불필요한 bounding box를 제거하고 검출 영역을 획득한다. 학습 및 테스트를 위해서 COCO Text dataset을 이용하였고 기울어진 텍스트에 대

하여 적절하게 회전된 영역을 획득할 수 있음을 확인하였다.

2. 본론

방향성을 고려한 텍스트 검출을 위한 네트워크 구조는 아래의 그림 1과 같다. CNN을 통해 고 수준의 특징 지도를 추출하고 RPN을 통해 영상 내 텍스트를 포함하는 bounding box를 예측하고 세 가지의 서로 다른 크기의 pooling된 특징 지도를 추출한다.

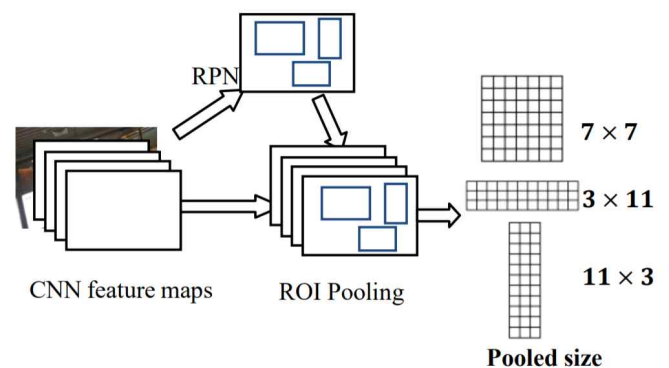


Figure 1. Architecture network for text detection

기존의 객체 검출에는 ROI Pooling 단계에서 7x7의 특징지도만을 추출하였지만 추가로 11x3, 3x11 크기의 특징지도를 추출하여 각각 수평과 수직 방향의 텍스트 검출을 용이하게 한다. Pooling을 통해 추출한 특징지도들을 병합하고 fully connected layer를 거쳐 텍스트와 텍

스트가 아닌 대상에 대한 score, 정렬된(axis-aligned) bounding box와 회전된 bounding box에 대한 좌표를 동시에 예측한다. 마지막으로 NMS를 통해 불필요한 잔여의 bounding box를 제거하고 텍스트 검출 결과를 획득한다.

회전된 bounding box 좌표 예측을 위한 학습데이터를 생성해야 하는데 아래 그림 2와 같이 정렬된 bounding box(녹색)를 회전된 bounding box(빨간색)으로 변환하고 네 위치의 코너 좌표를 획득한다. OpenCV의 minAreaRect 함수를 이용하여 생성하였다.

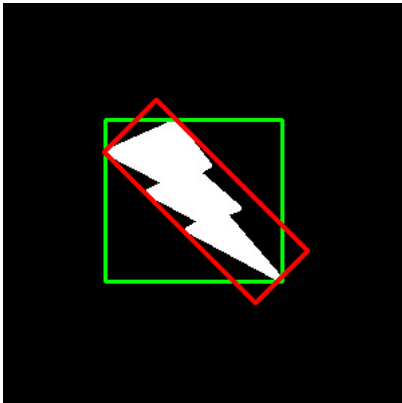


Figure 2. Normal bounding box and rotated bounding box

학습한 모델에 대하여 테스트 이미지를 입력하여 검출한 결과는 그림 3과 같다. 파란색 box는 ground truth이고 녹색 box가 텍스트 방향에 맞춰 회전된 영역이다.



Figure 3. Text detection results

3. 결론

본 논문에서는 Faster R-CNN 기반의 방향성을 고려한 장면 텍스

트 검출 기법을 제안한다. 텍스트의 방향 정보가 텍스트 인식이나 다른 응용 방법에 유용하기 때문에 축 정렬된 bounding box 뿐만 아니라 기울어진 텍스트에 맞춰 회전된 bounding box도 예측하도록 한다. 텍스트 검출이란 문제에 맞도록 세 가지의 다른 크기 및 비율의 특징지도를 추출하여 수평 및 수직 방향의 텍스트 검출 성능을 개선한다. 기존의 COCO Text 2017 dataset을 이용하였고 회전된 box에 대한 네 가지 꼭짓점 좌표(x, y)를 추출하여 학습데이터로 활용하였다. 학습한 모델에 테스트 이미지를 입력한 결과 기울어진 텍스트에 대하여 적절하게 회전된 bounding box를 예측하는 것을 확인하였다.

ACKNOWLEDGEMENT

본 논문은 2018년도 과학기술정보통신부 정보보호핵심원천기술개발사업의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임. (2018-0-00401, 영상 빅데이터 학습을 위한 De-identification 기술 개발)

참 고 문 헌

- [1] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks", In Advances in neural information processing systems, pp. 91-99, 2015
- [2] R. Gomez, B. Shi, L. Gomez, L. Numann, A. Veit, J. Matas and D. Karatzas, "ICDAR2017 robust reading challenge on COCO-Text", In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), IEEE, vol. 1, pp. 1435-1443, November, 2017
- [3] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016