

Visual Detection with Context for Document Layout Analysis

(2019)

김동원

목차

논문 개요

제안한 방법

Faster R-CNN

실험 & 결론

논문 개요

문서 레이아웃 분석을 통해 문서의 여러 문맥 정보를 제공하여 ocr와 문서 분류 등의 정확성을 높일 수 있다.

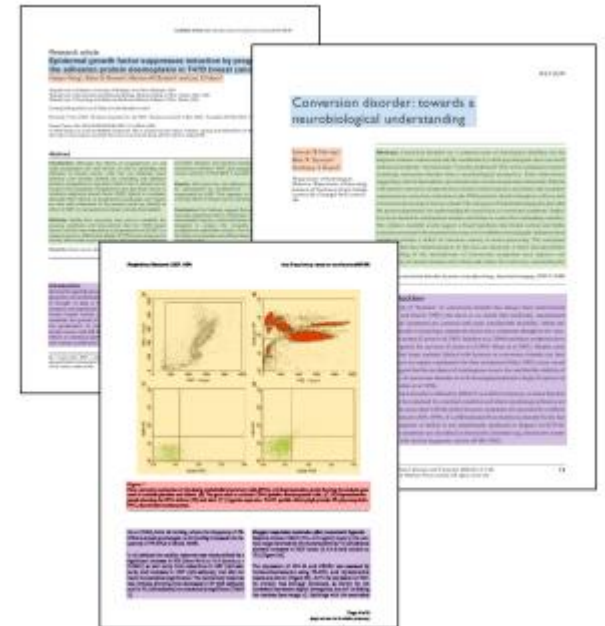
Faster R-CNN을 사용

이전 방법들

machine-learning based(rule-based) : 텍스트를 이용하여 text block에 라벨을 추정하여 부여하는 방법

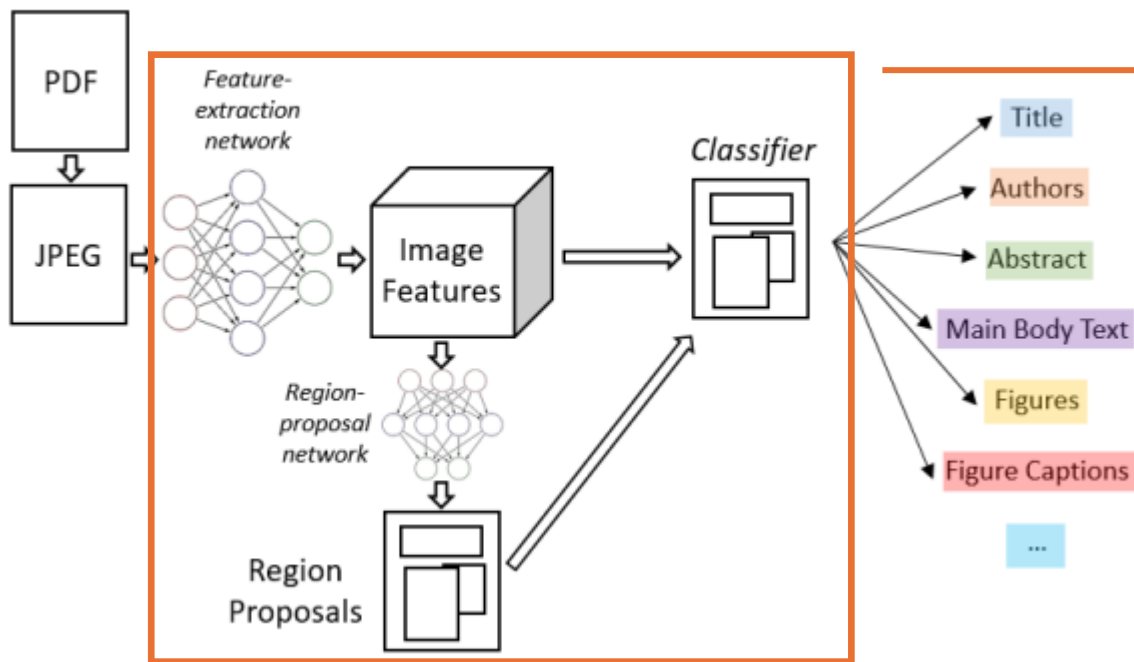
visual-based : text의 분할에 집중한 방법

이 논문은 context정보를 이용하여 detection을 수행한다.



제안한 방법

object detection 방법을 통해 문서의 구역을 인식하여 라벨을 부여한 방법으로 text특징을 이용하지 않아도 된다는 장점이 있다.

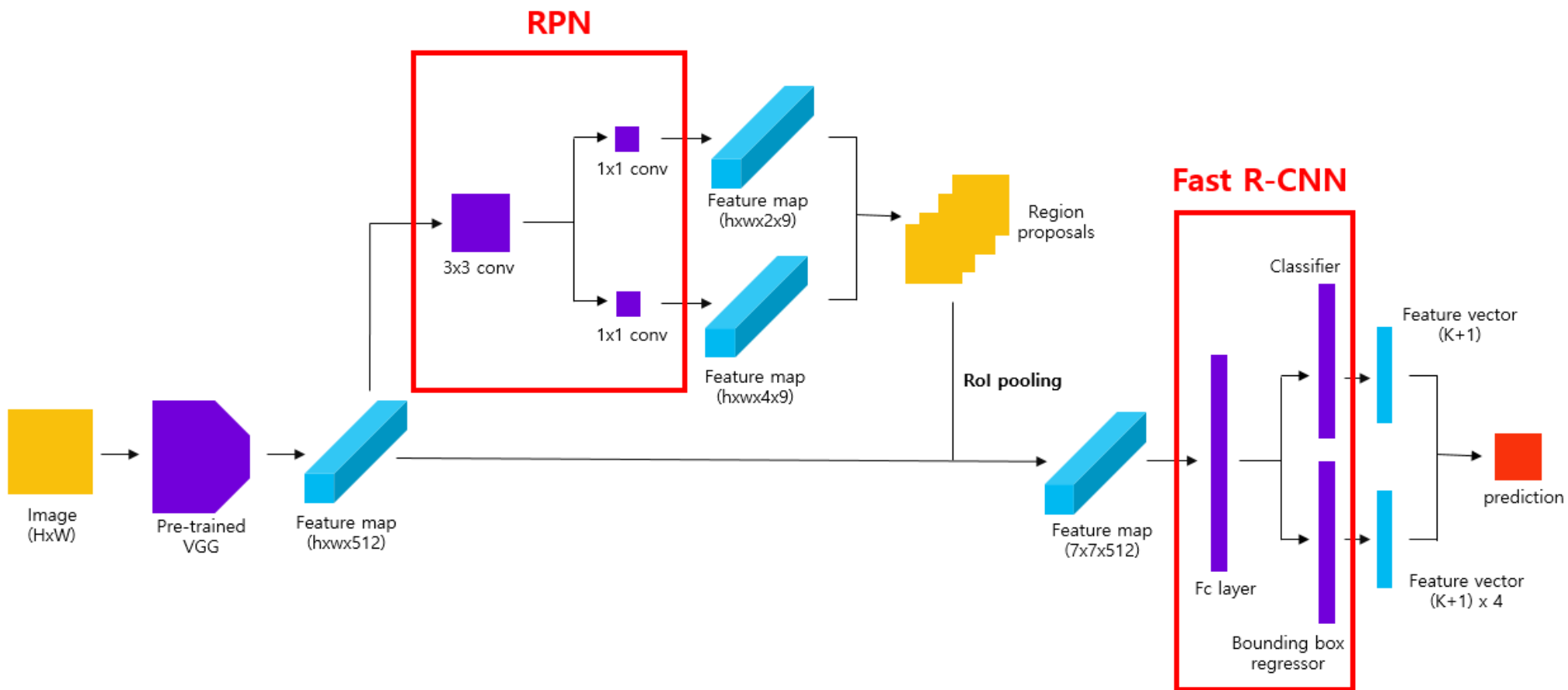


Faster R-CNN

1. RPN의 RoI를 분류(title, authors, abstract, ...)
2. context 정보를 제공하므로 분류 시 애매하거나 시각적으로 비슷한 것도 잘 분류할 수 있다.

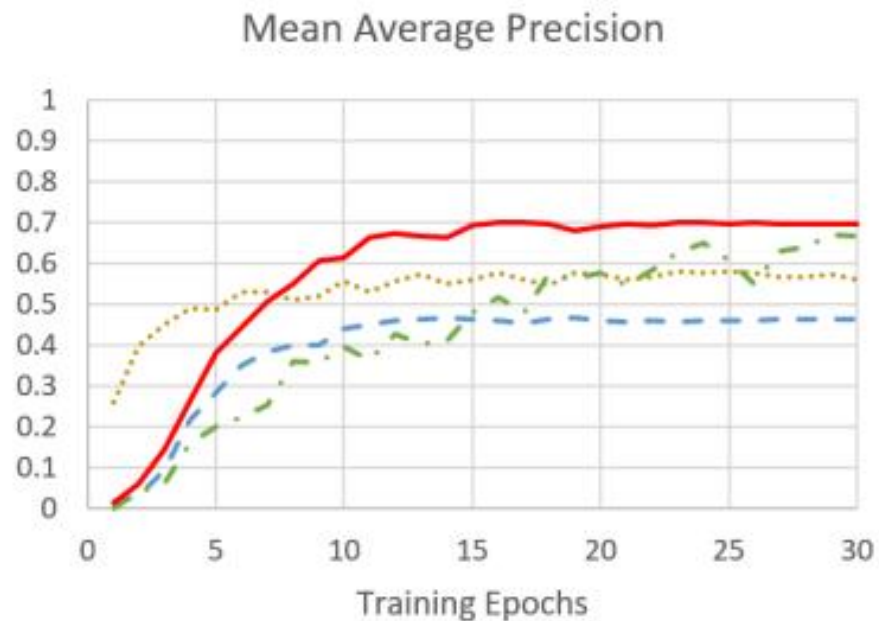
Faster R-CNN

fast r-cnn방식에서 selective search방식을 RPN(region proposal network)대체하므로 속도를 향상시킨 방법



실험 & 결론

base 모델 별 성능과, 분류 class별 성능



- Faster R-CNN (baseline)
- RetinaNet (reference)
- . - YOLOv3 (reference)
- Ours (includes contextual features)



- title
- body
- . - table
- authors
- figure
- table_caption
- abstract
- figure_caption
- references
- Mean Average Precision

context 정보가 분류 시 상당한 영향을 미치고, .scientific article의 경우 위치와 라벨이 매우 큰 상관관계를 가진다.

참고자료

[faster r cnn image](#)

Layout detection

yolov8

Class	Images	Instances	Box(P	R	mAP50	mAP50-95) :
all	57	456	0.815	0.794	0.852	0.684
Figure	21	33	0.83	0.879	0.886	0.731
List	2	3	0.54	0.667	0.637	0.515
Table	10	17	0.893	0.987	0.972	0.875
Text	55	298	0.861	0.87	0.931	0.777
Title	35	105	0.952	0.57	0.834	0.521