

OCR

[세나지] 1주차 발제자료



OCR

18기 분석 박규연

OCR

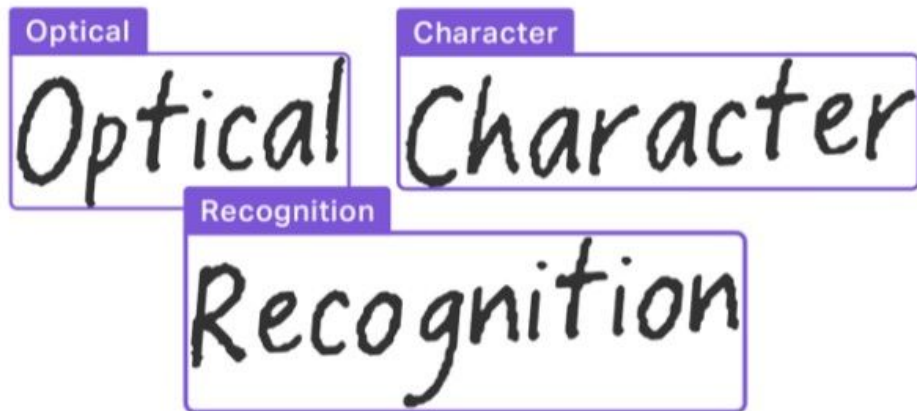
[세나지] 1주차 발제자료



OCR이란?

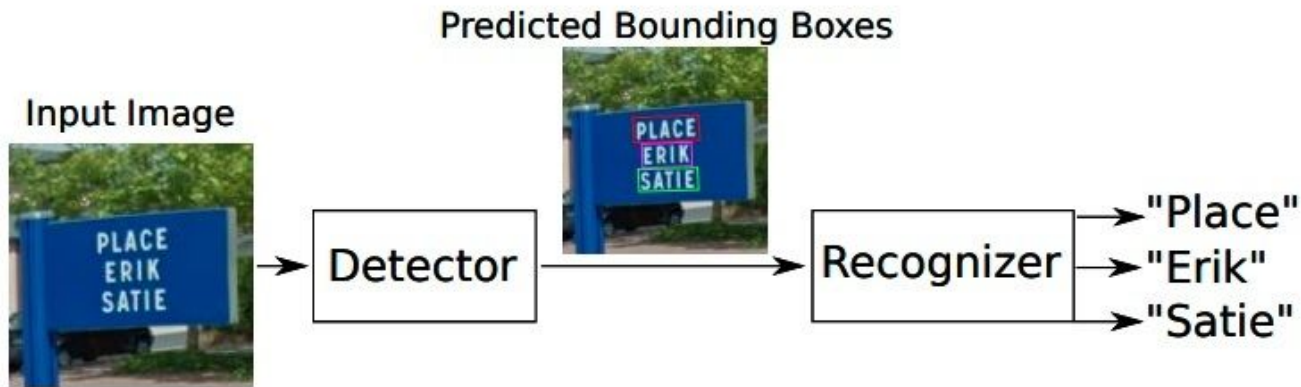
Optical Character Recognition

광학 문자 인식



OCR

[세나지] 1주차 발제자료

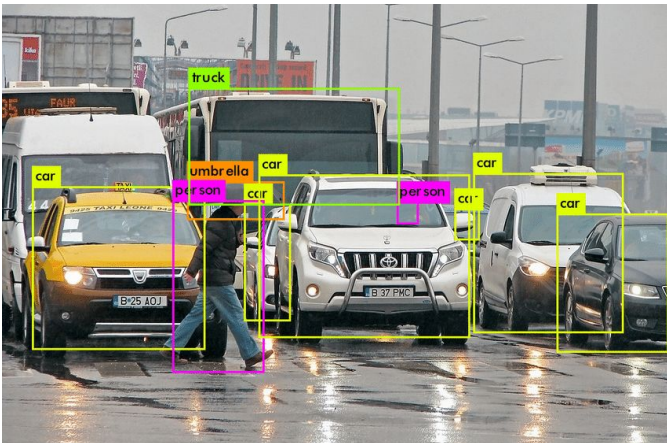


[출처 : <https://arxiv.org/pdf/1707.08831.pdf>]

OCR = Text detection + Text recognition

Text Detection

Object Detection(객체 탐지)을 Text 탐지로 확장한 개념



<https://wikidocs.net/142645>



https://docs.openvino.ai/latest/omz_models_model_text_detection_0004.html

Text Recognition

탐지한 **Text**가 어떤 문자들인지 인식해 내는 과정



<https://www.analyticsvidhya.com/blog/2020/05/build-your-own-ocr-google-tesseract-opencv/>

OCR

[세나지] 1주차 발제자료



딥러닝 이전 OCR

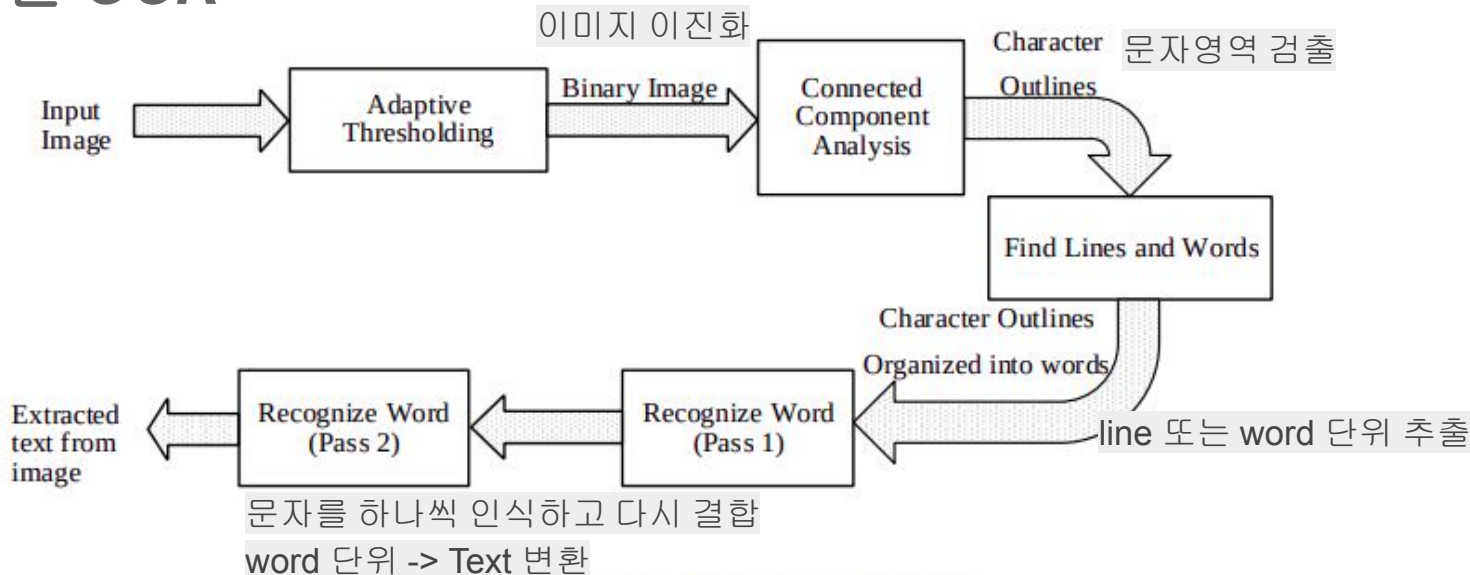


Fig. 4: Architecture of Tesseract OCR

[From Videos to URLs: A Multi-Browser Guide To Extract User's Behavior with Optical Character Recognition]

<https://arxiv.org/pdf/1811.06193.pdf>

Tesseract

2005년 오픈소스로 출시되어

2006년부터 구글의 후원을 받고 있는 오픈소스 OCR 엔진

최근에는 LSTM을 적용하는 등 딥러닝 베이스 아키텍처로 발전하고
있음

<https://github.com/tesseract-ocr/tesseract>



Tesseract OCR

딥러닝 적용 OCR

1. 우리가 원하는 단위로 문자 검출
2. 아키텍처 단순화
3. 검출, 인식을 동시에 해내는 End-to-End OCR 모델도 연구되고 있음

Text Detection

Text Detection

Object Detection, Segmentation

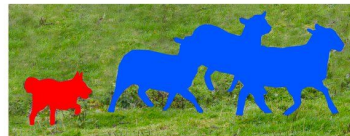
1. 문자 검출 최소 단위 (글자, 단어, 문장 단위 등)
2. 문장, 단어 단위 -> 길이 고려
3. 글자 단위 -> 단어, 문장으로 다시 묶어주는 과정

필요

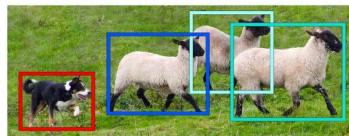
오늘 저녁은 엽떡이다.
내일 저녁은 소갈비다.
아 배고파.



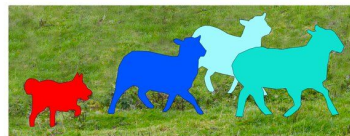
Image Recognition



Semantic Segmentation



Object Detection



Instance Segmentation

<https://wikidocs.net/148872>

Text Detection

1. 단어 단위 탐지

Object Detection의 **Regression**기반의 Detection 방법

Anchor를 정의하고 단어의 유무, Bounding Box의 크기를 추정해 단어를 찾아냄

이미지 상에 균일하게

Object가 존재하는 범위를 박스 단위로

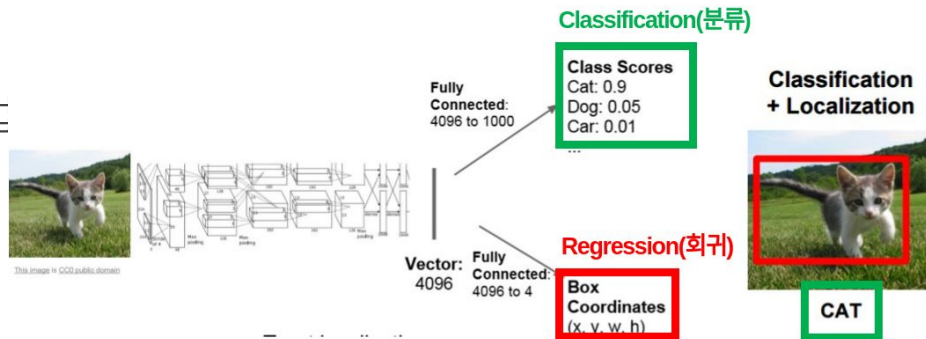
2. ^{박스처리}글자 단위 탐지

표기

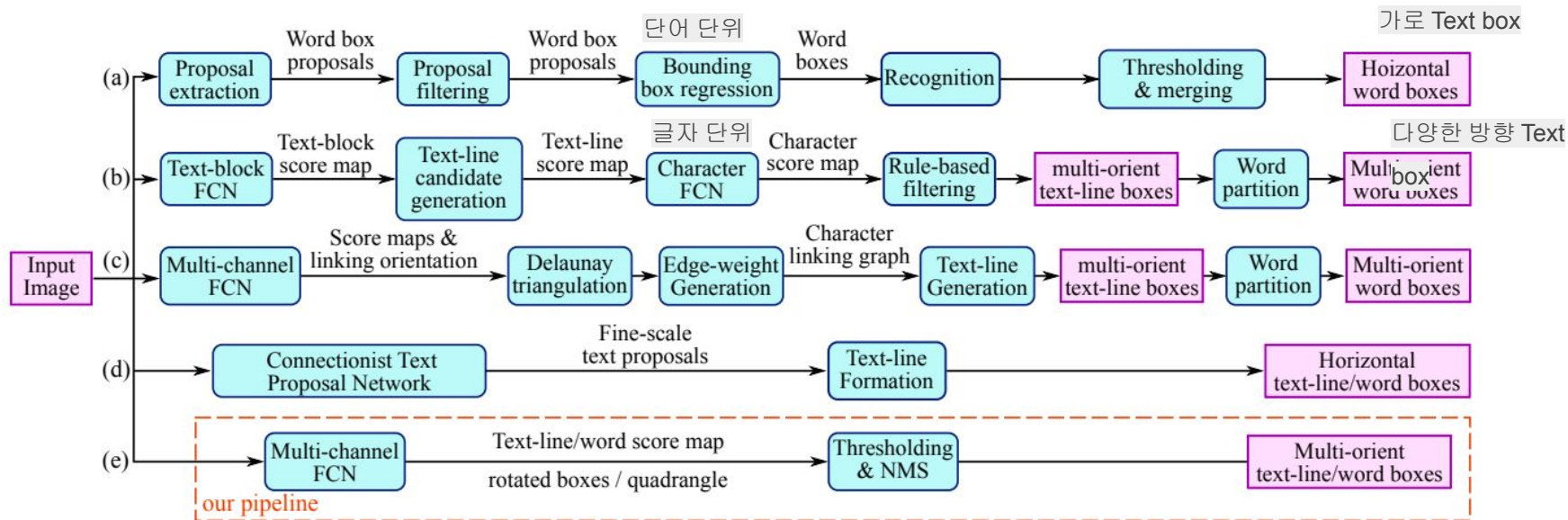
글자 영역을 **Segmentation**하는 방법

Object가 존재하는 영역 분할

3. 둘 다 이용

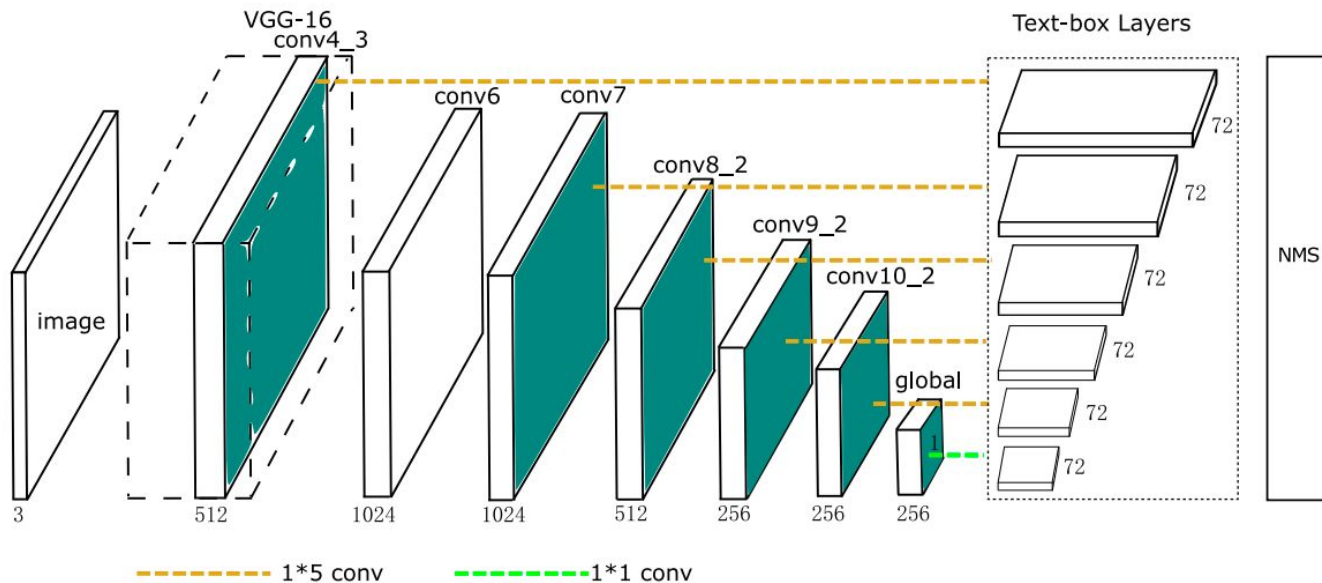


Text Detection에서 Bounding Box의 활용



Regression - TextBoxes

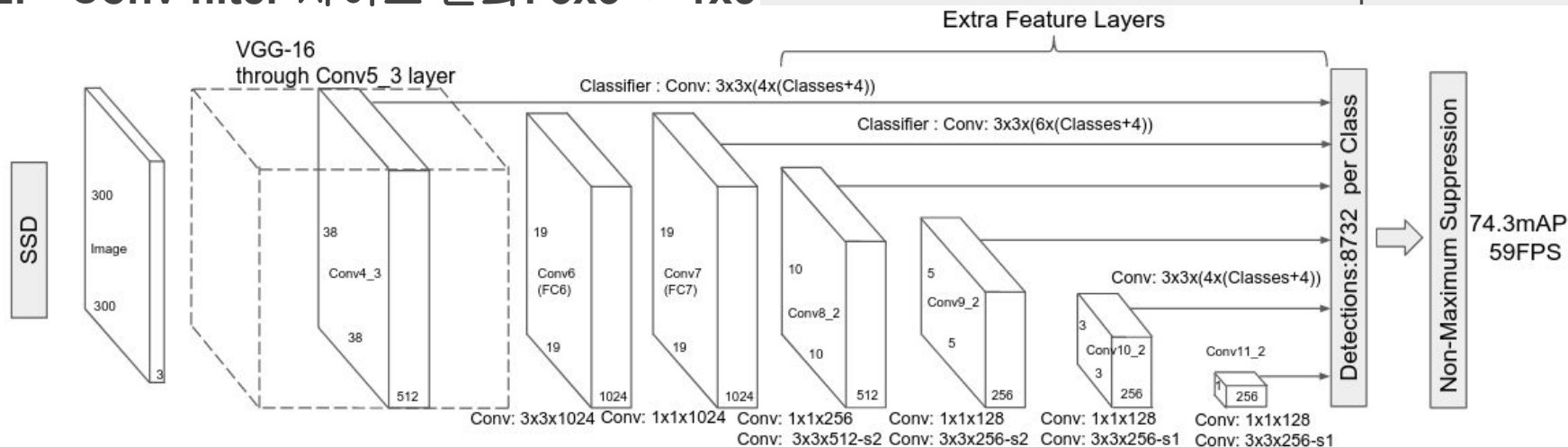
딥러닝 기반의 Detection 모델을 활용해서 단어 단위로 인식



1. 네트워크 기본 구조: SSD(Single Shot Multibox Detector)

2. Conv filter 사이즈 변화: 3x3 -> 1x5

일반적으로 단어들은 가로로 길기 때문에 Aspect Ratio가 큼



[SSD: Single Shot Multibox Detector] <https://arxiv.org/pdf/1512.02325.pdf>

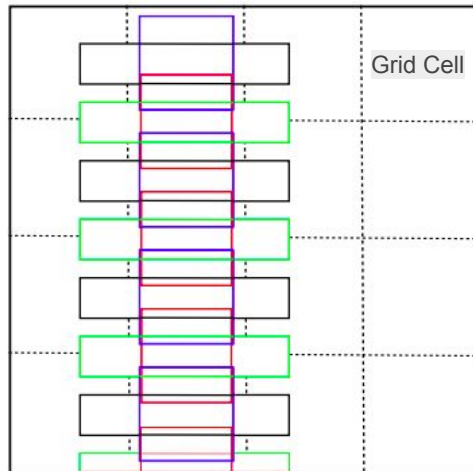
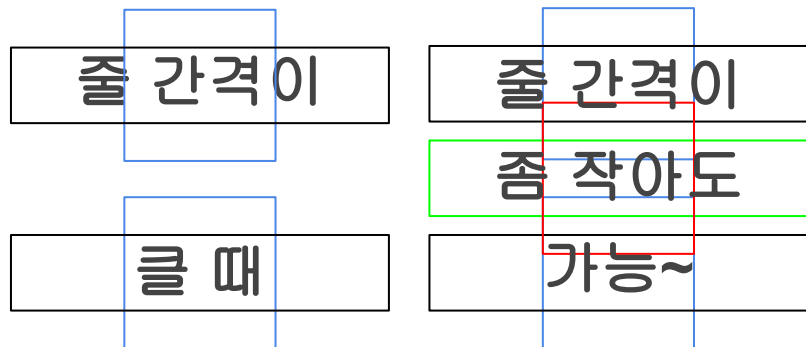


Figure 2: Illustration of default boxes for a 4*4 grid. For better visualization, only a column of default boxes whose aspect ratios 1 and 5 are plotted. The rest of the aspect ratios are 2,3,7 and 10, which are placed similarly. The black (aspect ratio: 5) and blue (ar: 1) default boxes are centered in their cells. The green (ar: 5) and red (ar: 1) boxes have the same aspect ratios and a vertical offset(half of the height of the cell) to the grid center respectively.

3. Anchor box를 ^{1, 2, 3, 5, 7} 큰 Aspect Ratio로 만들고 여기에 vertical offset 적용

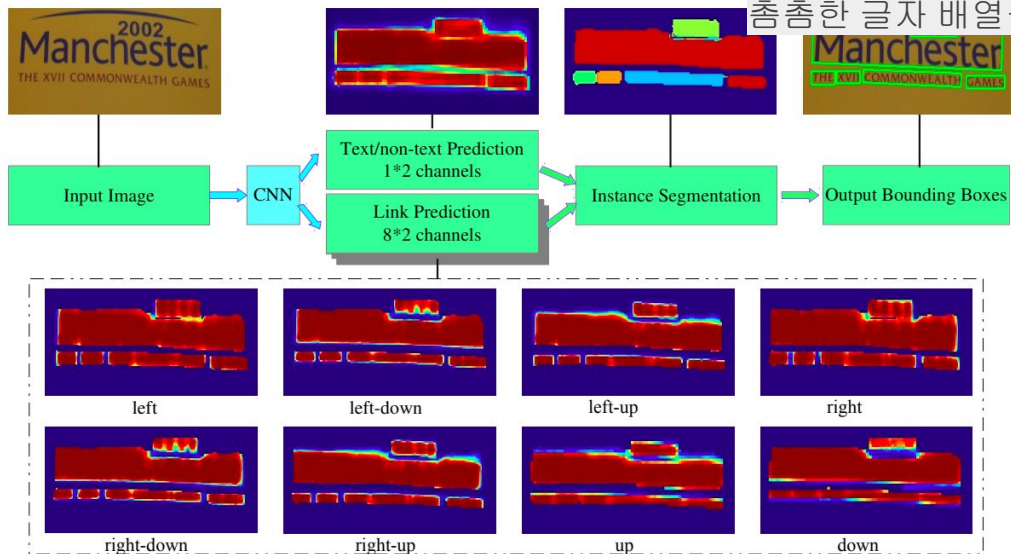
세로 방향으로 촘촘한 단어 배열에 대응



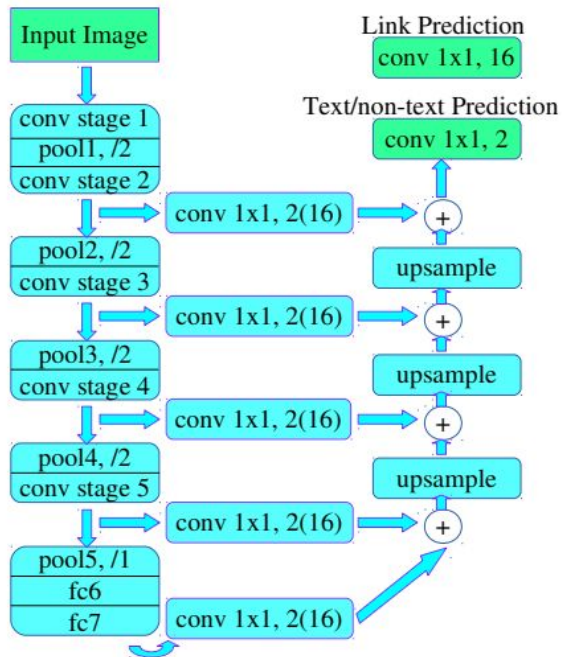
Segmentation - PixelLink

Segmentation을 이용해서 배경과 글자가 있는 영역을 분리할 수
있음

쫄쫄한 글자 배열들을 쪼개고 의미있는 단위로 다시 연결해줘야 함



1. 글자 영역과 아닌 영역 분리
2. 연결되는 방향 정보



Output

- **Text/non-text Prediction (1*2):** 글자 영역 여부
- **Link Prediction (8*2):** 인접한 8개의 Pixel에 대한 연결 여부

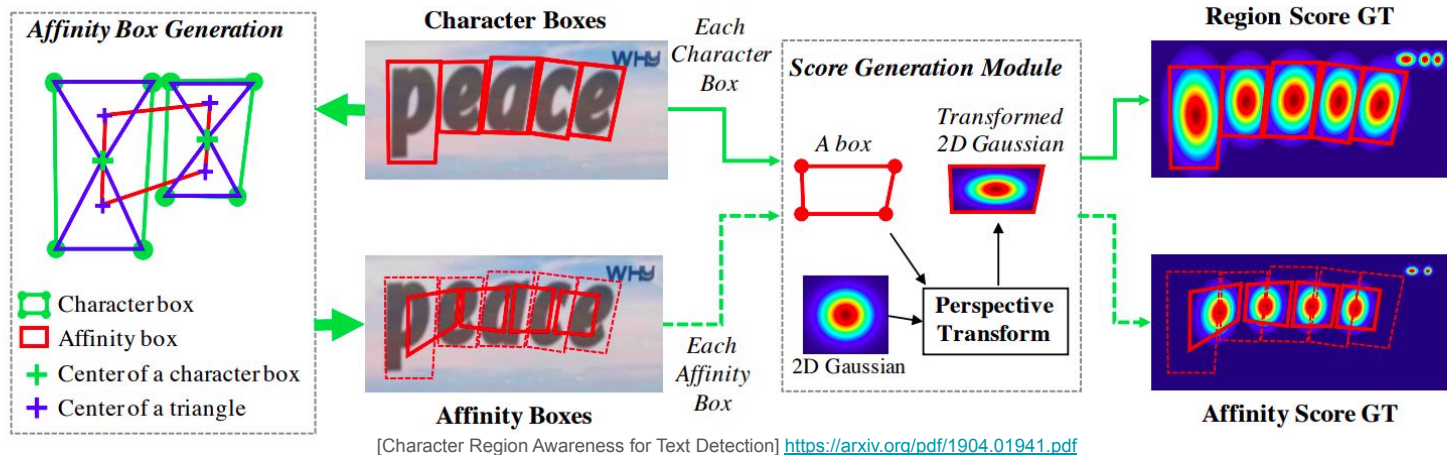
인접한 pixel이 중심 pixel과

단어 단위로 연결된 pixel인지 / 분리된 pixel인지 알 수 있음

-> 단어 단위의 instance segmentation 가능

CRAFT

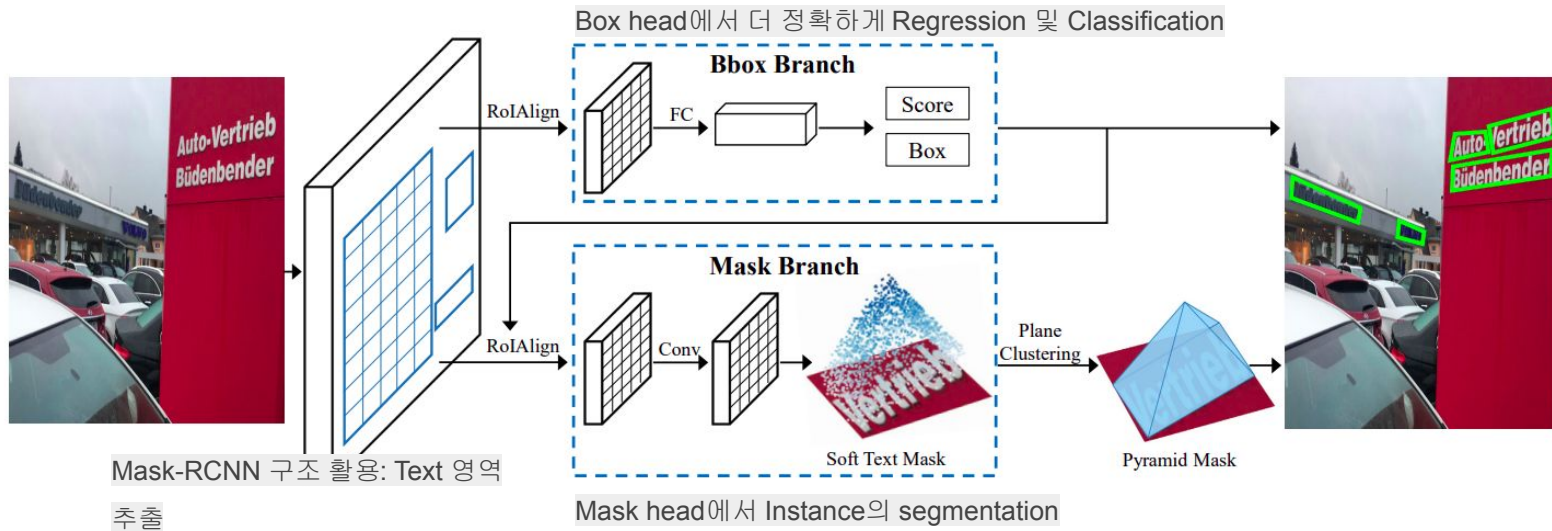
Character 단위로 문자의 위치를 찾아낸 뒤 이를 연결하는 방식을 Segmentation 기반으로 구현



글자의 영역을 명확하게 구분하지 않고

가우시안 분포를 따르는 원형의 score map을 만들어서 배치시키는 방법으로 문자의 영역을

PMTD (Pyramid Mask Text Detector)

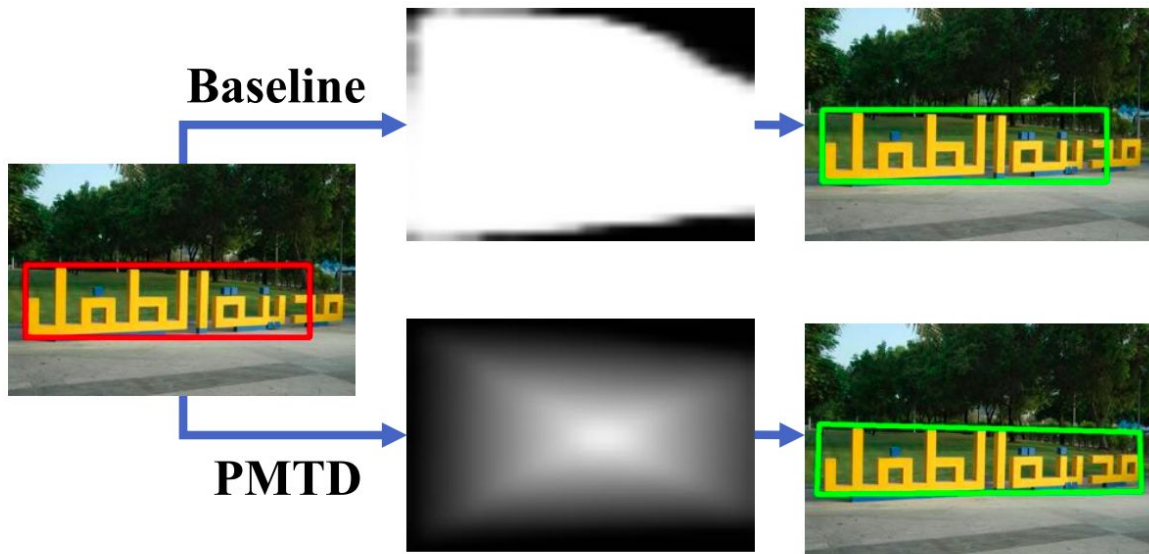


OCR

[세나지] 1주차 발제자료



Mask-RCNN: boundary를 모두 text 영역으로 잡음



PMTD: 단어의 사각형 배치 특성을 반영하 피라미드 형태의 score map

활용

[Pyramid Mask Text Detector] <https://arxiv.org/pdf/1903.11800.pdf>

-> 피라미드 형태의 mask 생성

OCR

[세나지] 1주차 발제자료



Text Recognition

Unsegmented Data

적절히 **segment**가 되어있어야 문자 인식에 용이하지만,

분리 비용이 비싸기 때문에 데이터가 적절히 **segment** 되어 있지 않는 경우가 많음.

segment 되어 있지 않은 하위 데이터들끼리 **시퀀스(sequence)**를 이루고 있음.

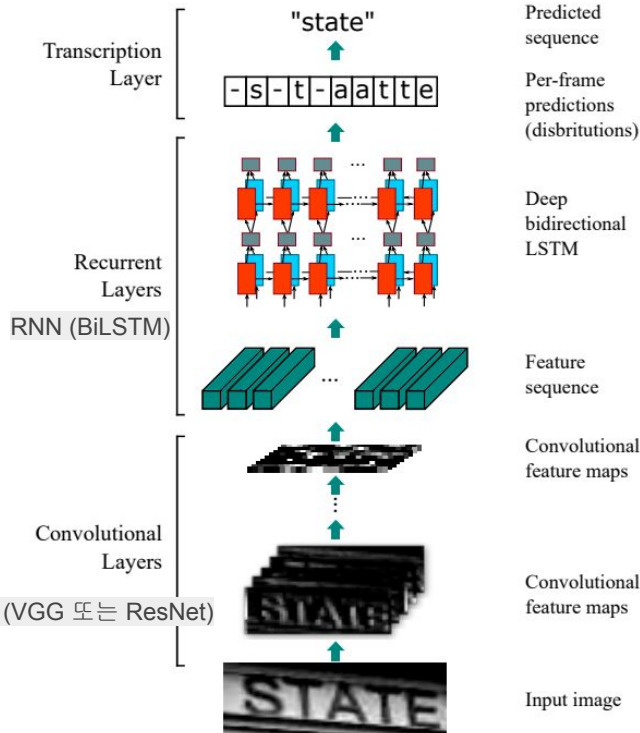
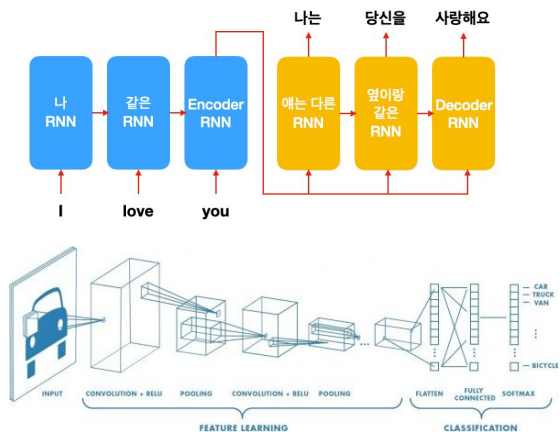
글자 이미지들은 문자의 순서대로 정보를 가지고
있음

OCR

[세나지] 1주차 발제자료



CRNN (CNN + RNN)



Predicted sequence

Per-frame predictions (disbtrutions)

Deep bidirectional LSTM

Feature sequence

Convolutional feature maps

Convolutional feature maps

Input image

sequence의 앞뒤 정보를
이용하여 문자 예측

sequence 형태의 feature로 변환

문자의 정보를 가진 feature map
추출

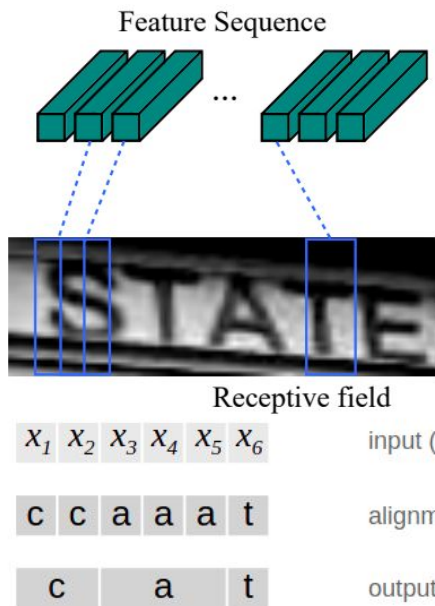
[An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition]

<https://arxiv.org/pdf/1507.05717.pdf>

CTC (Connectionist Temporal Classification)

클래스 라벨 순서만 있고

각 클래스 위치는 알 수 없을 때



TPS (Thin Plate Spline Transformation)

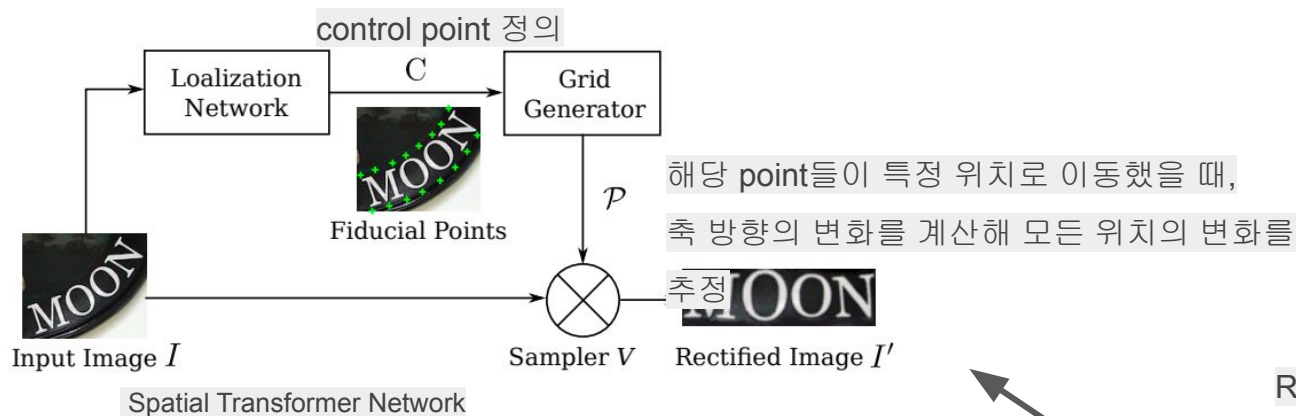
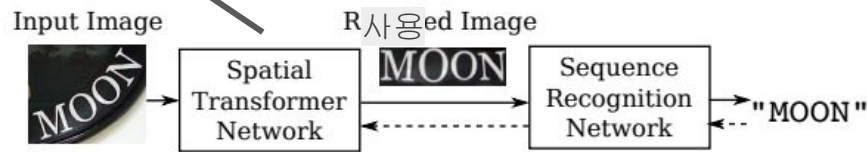


Figure 2. **Structure of the STN**. The localization network localizes a set of fiducial points C , with which the grid generator generates a sampling grid \mathcal{P} . The sampler produces a rectified image I' , given I and \mathcal{P} .

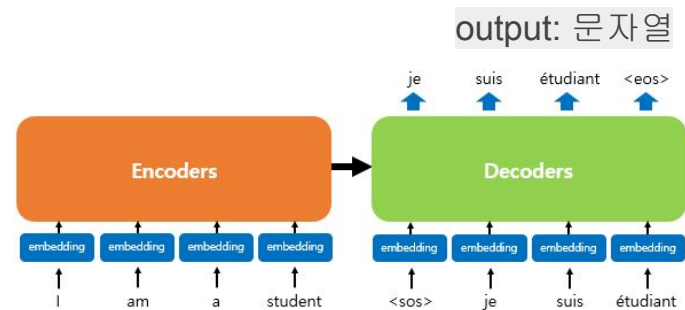
Recognition Network 앞에 붙여
이미지를 정방향으로 맞춰줄 때



Transformer

언어 처리 분야의 언어 번역을 위해 등장한 모델 구조

attention mechanism 사용



input: 문자열

<https://wikidocs.net/31379>

Encoder

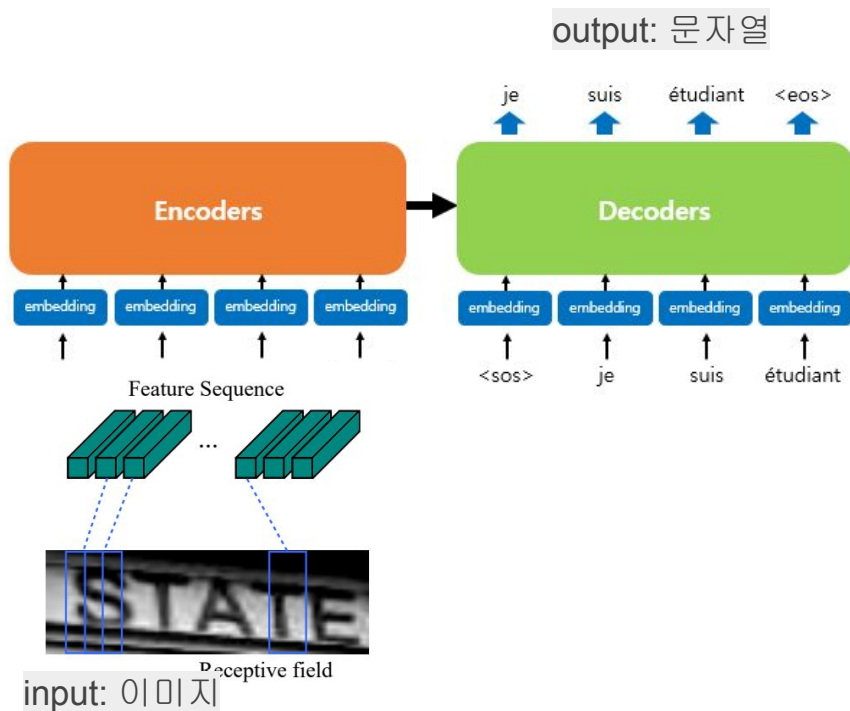
- 토큰들 간의 관계를 추론해내면서 **attention** 추출
집중해야 될 부분!!

Decoder

- 그 다음에 나올 글자를 추론

OCR

[세나지] 1주차 발제자료



Encoder

- 입력된 이미지의 특징 추출
- 이미지 내 픽셀 별 관계 추출

Decoder

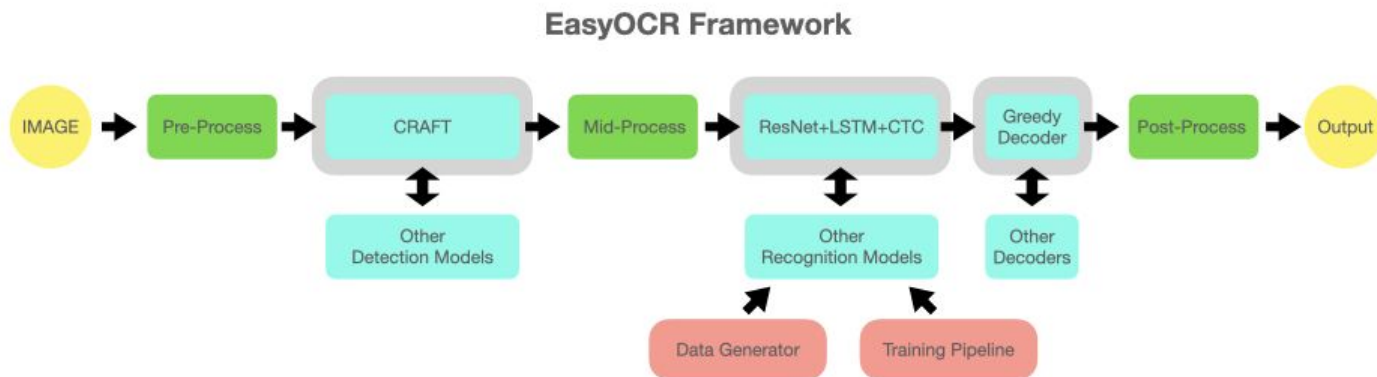
- 실제 문자열 글자 간 관계 추출
- 이미지와 실제 문자열 간 관계 추출

OCR

[세나지] 1주차 발제자료



EasyOCR



<https://github.com/JaidedAI/EasyOCR>

Text Detection, Text Recognition 관련 논문 정리 Repository

hwalsuklee/**awesome-deep-text-detection-...**



A curated list of resources for text detection/recognition (optical character recognition) with deep learning methods.

5

Contributors

9

Issues

2k

Stars

519

Forks



<https://github.com/hwalsuklee/awesome-deep-text-detection-recognition>

References

<https://github.com/hwalsuklee/awesome-deep-text-detection-recognition>

https://www.youtube.com/watch?v=hn26j_QOIXk

<https://ehdgns.tistory.com/126>

OCR

[세나지] 1주차 발제자료



감사합니다