DEVIEW
2019

문자인식(OCR), 얼마나 정확하지?
(문자인식 성능을 정확하게 측정하는 방법)

최찬규 (chankyu.choi@navercorp.com)
파파고, 이미지 번역팀

NAVER

CONTENTS

DEVIEW
2019

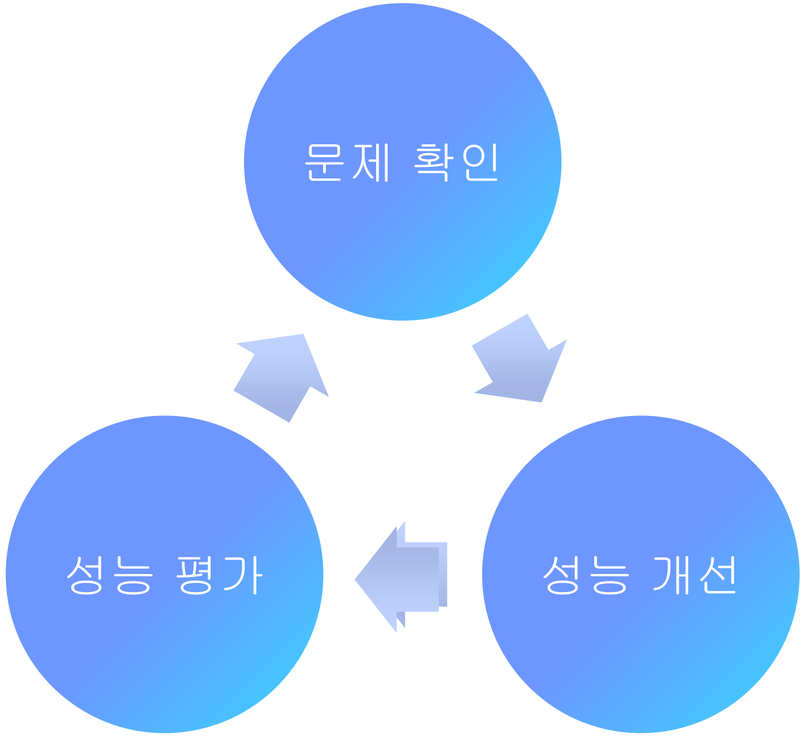
1.성능 평가의 중요성
(왜 성능 평가인가?)

1.1 성능 평가의 중요성

DEVIEW 2019

1) 성능을 개선하기 위해!!

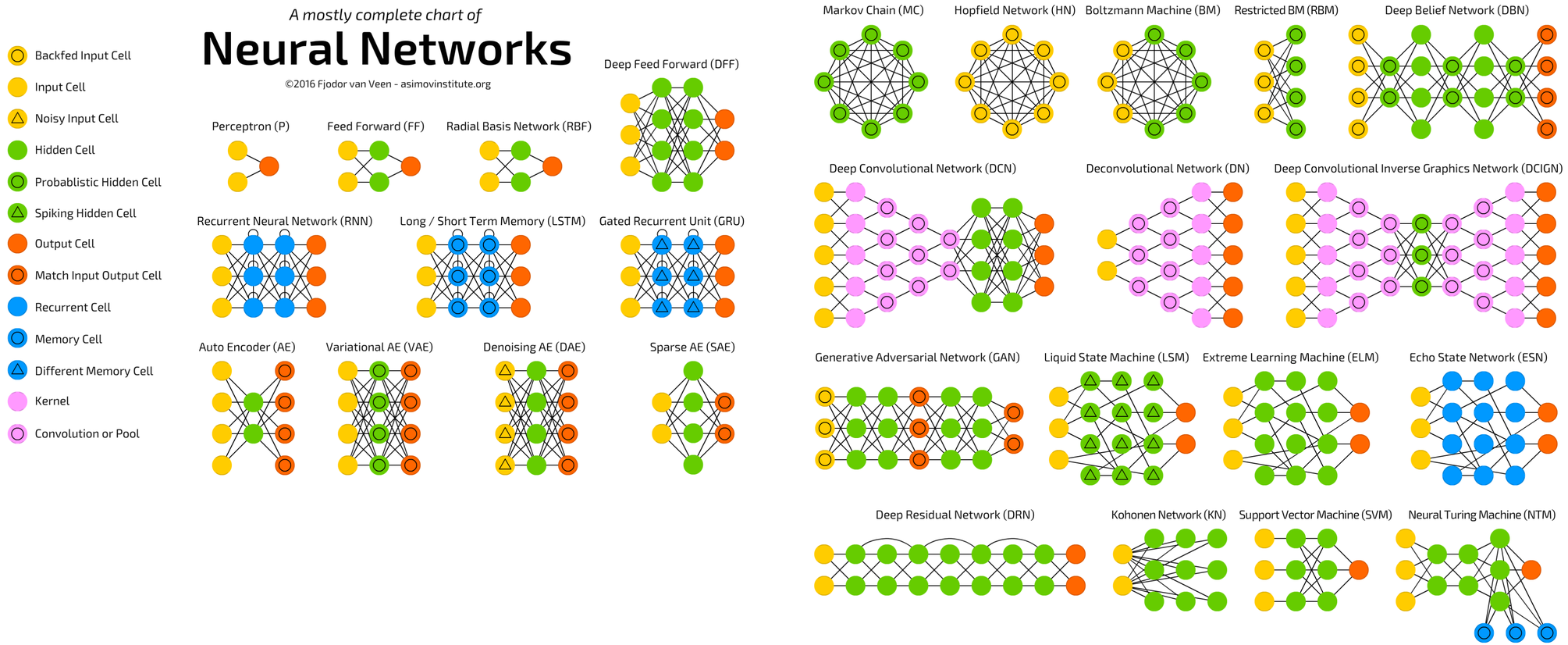
- 성능을 알아야, 문제점을 파악할 수 있고, 성능을 개선할 수 있음



1.1 성능 평가의 중요성

DEVIEW 2019

2) 모델 선택 : 많은 모델 중, 어떤 모델을 서비스 할 것인가?



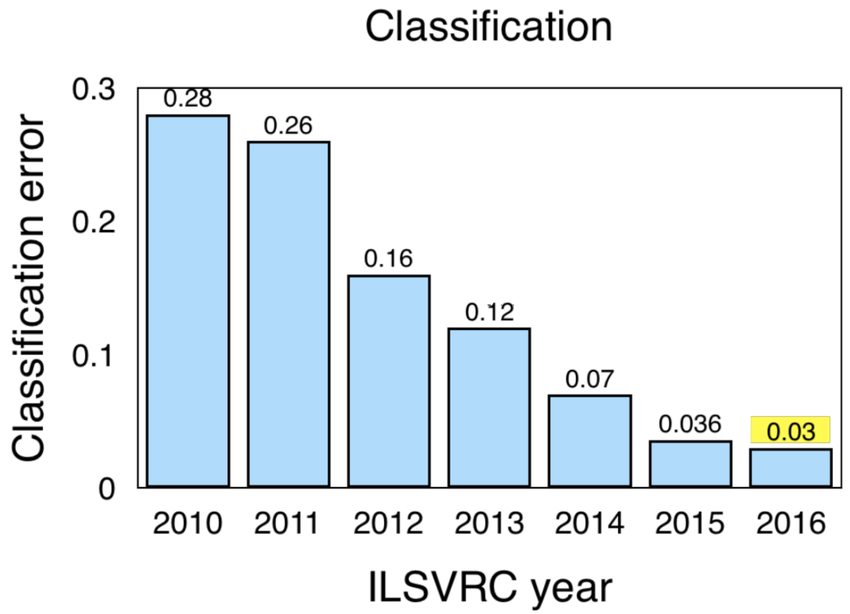
이미지 출처: {http://www.asimovinstitute.org/wp-content/uploads/2016/09/neuralnetworks.png}

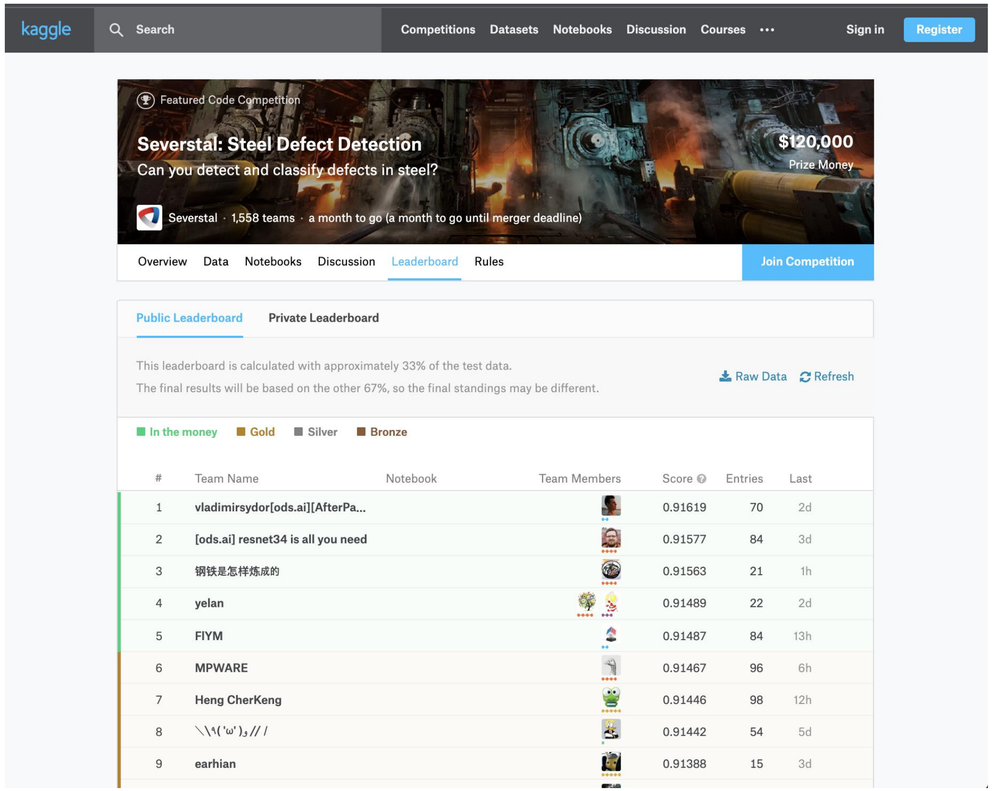
1.1 성능 평가의 중요성

DEVIEW 2019

3) 다른 연구 그룹과의 성능 비교

- 기술의 진보를 가져다 준 ImageNet(ILSVRC)과 Kaggle 경진대회



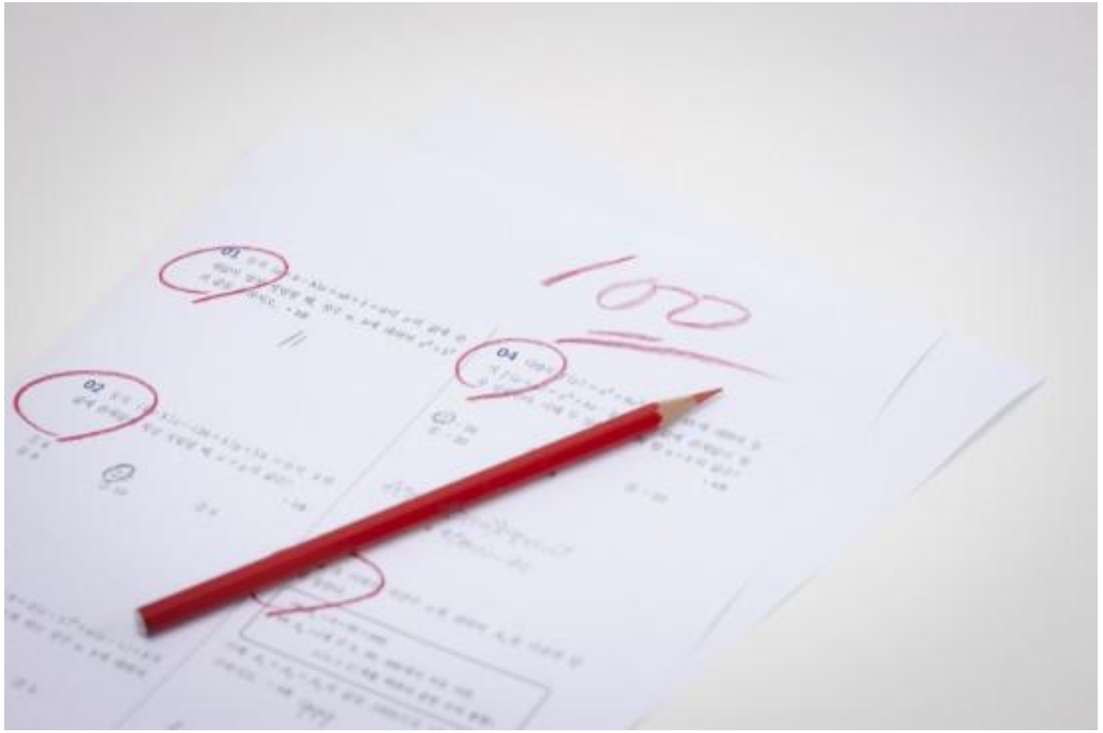


1.2 성능평가 어떻게 하고 계셔요?

DEVIEW
2019

1) 눈으로? (사람이 직접 평가)

- 시간과 비용도 많이 들고, 평가에 오류가 있을 수 있음

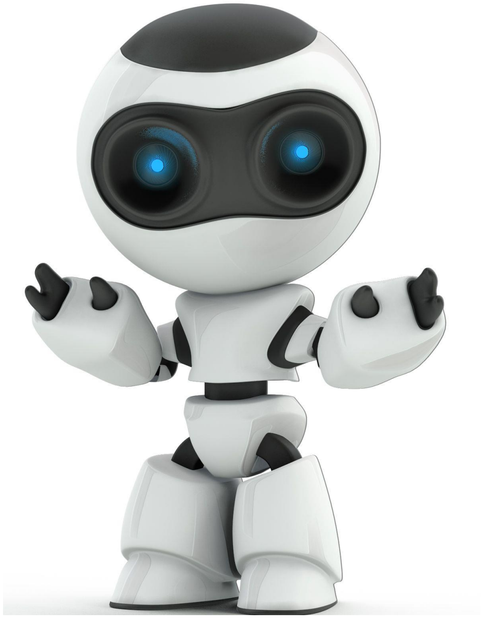


이미지 출처: {https://mblogthumb-phinf.pstatic.net/20141021_135/kais1995_1413854662101gafas_JPEG/tis034e14070014.jpg?type=w2}

1.2 성능평가 어떻게 하고 계셔요?

2) 컴퓨터? (자동 평가 시스템)

- 시간과 비용이 거의 들지 않으며, 평가에 오류가 없음 (오늘의 발표 내용)



이미지 출처: {https://www.pinterest.com.au/pin/417286721717116858/}

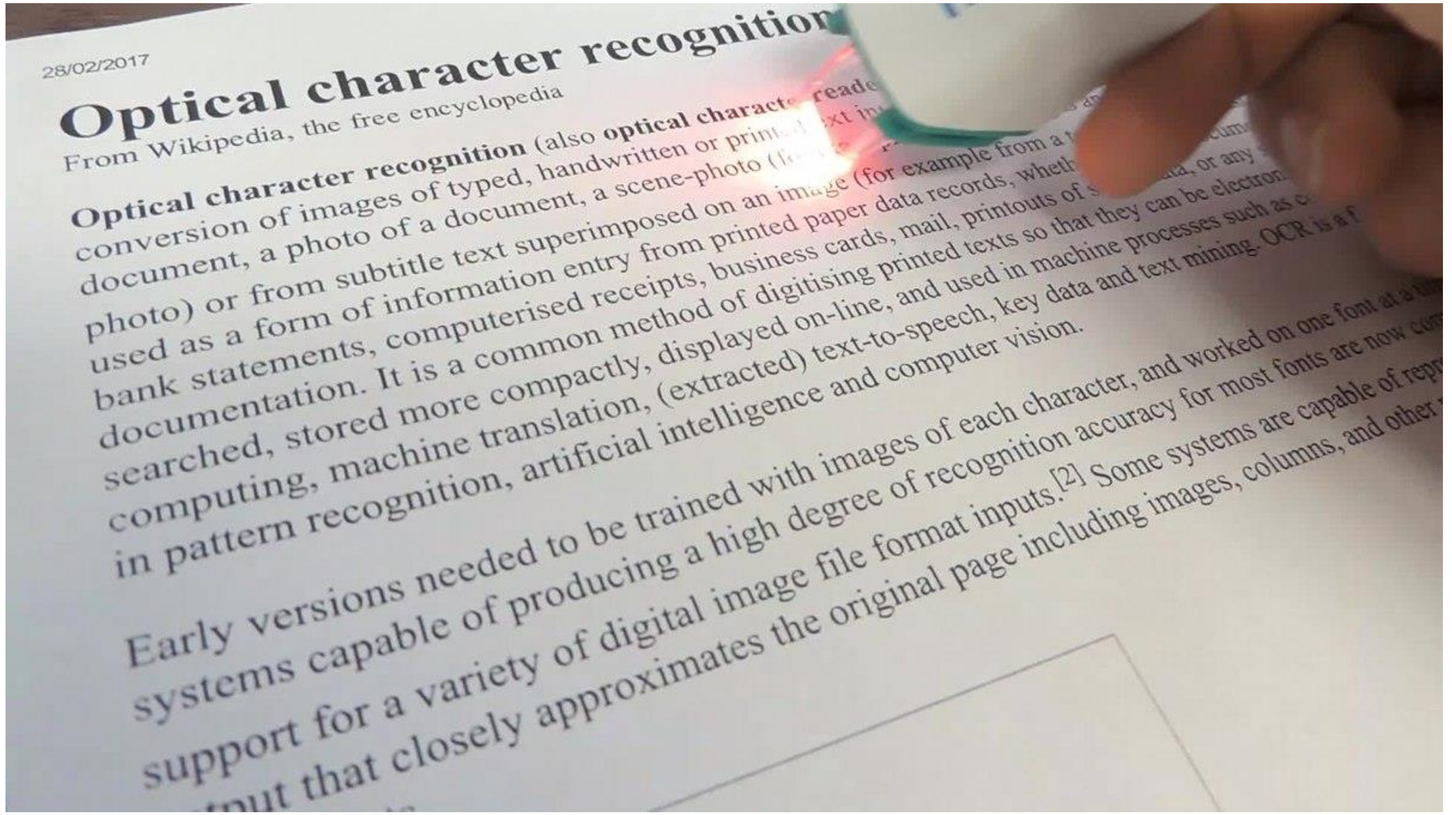
DEVIEW
2019

2.문자인식 개론

2.1 문자인식(Optical Character Recognition)

DEVIEW 2019

오프라인의 글자를 기계가 읽을 수 있도록 디지털화 하는 것



이미지 출처: {https://en.wikipedia.org/wiki/Optical_character_recognition}

2.2 문자인식 응용분야

DEVIEW
2019

자동차 번호판/명함/신용카드/신분증 인식, 이미지 번역

2.3 문자인식 과정

DEVIEW
2019

문자 검출 (Text Detection) → 문자 인식

2.3 문자인식 과정

DEVIEW
2019

문자 검출 → 문자 인식(Text Recognition)

“RIVERSIDE”

“WALK”

3.기존 평가 방법

3.1 사전지식 (Recall, Precision)

DEVIEW 2019

암 진단에 대한 평가 (경우의 수)

		의사의 진단 (예측)	
		True (양성)	False (음성)
실제 암의 유무	True (양성)	True Positive	False Negative
	False (음성)	False Positive	True Negative

이미지 출처: {https://www.ownyourhealthwa.org/wp-content/uploads/HowToShareYourOwnResearchWithYourDoctor-1024x943.jpg}

3.1 사전지식 (Recall, Precision)

DEVIEW 2019

암 진단에 대한 평가 (확률)

		의사의 진단 (예측)	
		True (양성)	False (음성)
실제 암의 유무	True (양성)	True Positive	False Negative
	False (음성)	False Positive	True Negative

Recall : 실제 True 중에서, True로 올바르게 예측한 비율 ($= \frac{TP}{TP + FN}$)

이미지 출처: {https://www.ownyourhealthwa.org/wp-content/uploads/HowToShareYourOwnResearchWithYourDoctor-1024x943.jpg}

3.1 사전지식 (Recall, Precision)

DEVIEW 2019

암 진단에 대한 평가 (확률)

		의사의 진단 (예측)	
		True (양성)	False (음성)
실제 암의 유무	True (양성)	**True** Positive	**False** Negative
	False (음성)	**False** Positive	**True** Negative

Recall : 실제 True 중에서, True로 올바르게 예측한 비율 ($= \frac{TP}{TP + FN}$)

Precision : 예측한 True 중에서, True로 올바르게 예측한 비율 ($= \frac{TP}{TP + FP}$)

이미지 출처: {https://www.ownyourhealthwa.org/wp-content/uploads/HowToShareYourOwnResearchWithYourDoctor-1024x943.jpg}

3.1 사전지식 (Recall, Precision)

DEVIEW 2019

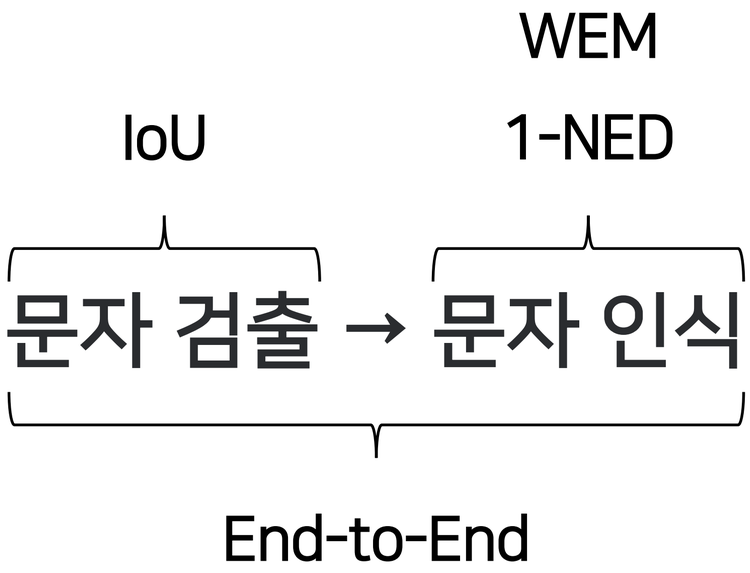
암 진단에 대한 평가 (확률)

		의사의 진단 (예측)	
		True (양성)	False (음성)
실제 암의 유무	True (양성)	True Positive	False Negative
	False (음성)	False Positive	True Negative

- Recall : 실제 True 중에서, True로 올바르게 예측한 비율 ($= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$)
- Precision : 예측한 True 중에서, True로 올바르게 예측한 비율 ($= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$)
- Recall과 Precision은 **반비례 관계**에 있어 **두 지표를 같이 봐야 함** ($H_{mean} = \frac{2 \cdot \mathrm{R} \cdot \mathrm{P}}{\mathrm{R} + \mathrm{P}}$)

이미지 출처: {https://www.ownyourhealthwa.org/wp-content/uploads/HowToShareYourOwnResearchWithYourDoctor-1024x943.jpg}

3.2 기존 평가 방법



3.2 기존 평가 방법

검출 평가 방법 (IoU : Intersection over Union)

- 정답(Ground Truth)과 예측(Prediction) 박스가 얼마나 겹치는지 확인 (50% 기준)

GT

Pred

$$IoU = \frac{R_{Pred} \cap R_{GT}}{R_{Pred} \cup R_{GT}}$$

$$= \frac{\text{겹친 영역}}{\text{합친 영역}}$$

$$= 0.72 \geq \mathbf{0.5}$$

이미지 출처: {http://host.robots.ox.ac.uk/pascal/VOC/}

3.2 기존 평가 방법

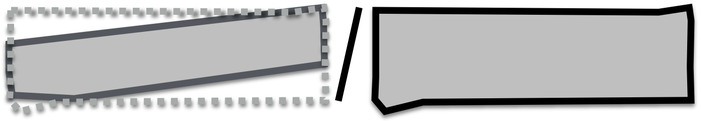
DEVIEW 2019

검출 평가 방법 (IoU : Intersection over Union)

- 정답(Ground Truth)과 예측(Prediction) 박스가 얼마나 겹치는지 확인 (50% 기준)

— GT
— Pred

$$IoU = \frac{R_{Pred} \cap R_{GT}}{R_{Pred} \cup R_{GT}}$$

$$= \text{  }$$

$$= 0.74 \geq \mathbf{0.5}$$

이미지 출처: {https://rrc.cvc.uab.es/?ch=2&com=evaluation&view=method_sample&task=4&m=3811&file=1&eval=1&sample=41&gtv=10}

3.2 기존 평가 방법

DEVIEW 2019

검출 평가 방법 (IoU : Intersection over Union)

- 정답(Ground Truth)과 예측(Prediction) 박스가 얼마나 겹치는지 확인 (50% 기준)

— GT
— Pred

$$IoU = \frac{R_{Pred} \cap R_{GT}}{R_{Pred} \cup R_{GT}}$$

$$= \quad /$$

$$= 0.65 \geq \mathbf{0.5}$$

이미지 출처: {https://rrc.cvc.uab.es/?ch=2&com=evaluation&view=method_sample&task=4&m=3811&file=1&eval=1&sample=41&gtv=10}

3.2 기존 평가 방법

DEVIEW 2019

인식 평가 방법 (WEM : Word based Exactly Matching)

- 정답과 예측 단어가 정확히 일치하는지 체크 (단어기반)

WEM("RIVERSIDE", "BIVERSIDE") = 0.00

WEM("WALK", "WALK") = 1.00

이미지 출처: {https://rrc.cvc.uab.es/?ch=2&com=evaluation&view=method_sample&task=4&m=3811&file=1&eval=1&sample=41&gtv=10}

3.2 기존 평가 방법

인식 평가 방법 (1-NED : Normalized Edit Distance)

- 두 단어간 편집 거리(삽입, 수정, 삭제)를 측정한 뒤, 긴 단어의 길이로 정규화 (글자기반)

$$1-NED(\text{“RIVERSIDE”}, \text{“BIVERSIDE”}) = 1 - \frac{1}{9} = 0.89$$

$$1-NED(\text{“WALK”}, \text{“WALK”}) = 1 - \frac{0}{4} = 1.00$$

3.2 기존 평가 방법

DEVIEW
2019

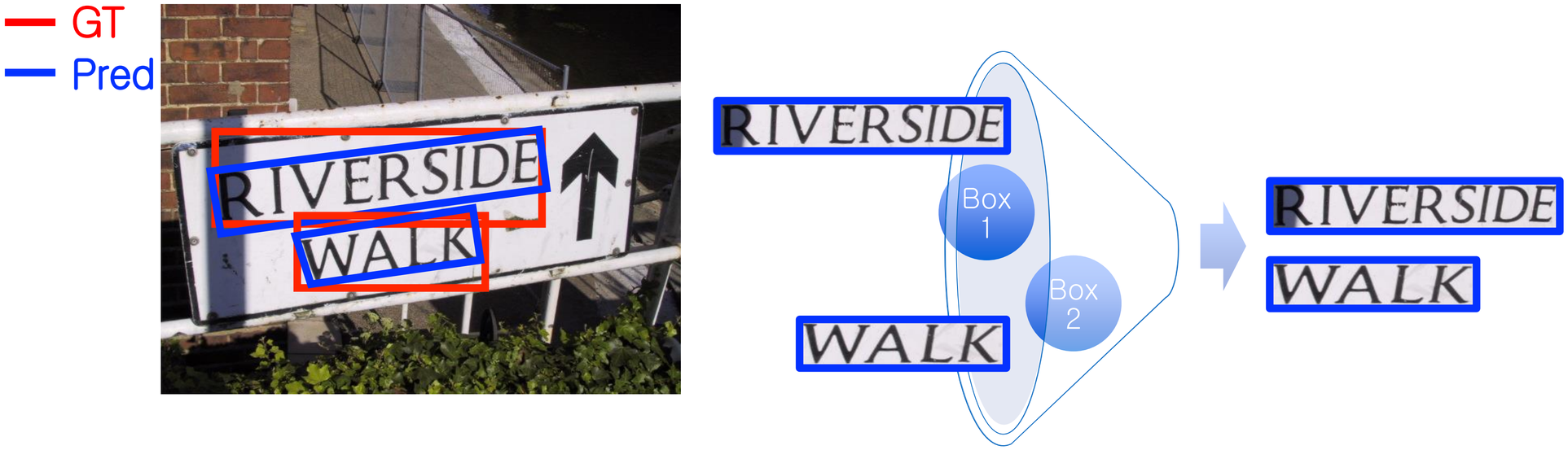
End-to-End(검출+인식) 평가 방법

- 순차(Cascade) 평가 처리 : 검출 평가(IoU) → 인식 평가(WEM, 1-NED)

3.2 기존 평가 방법

End-to-End(검출+인식) 평가 방법

- 순차(Cascade) 평가 처리 : 검출 평가(IoU) → 인식 평가(WEM, 1-NED)

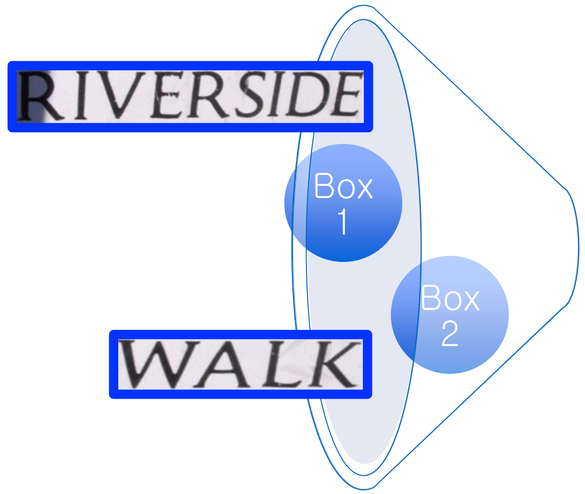


3.2 기존 평가 방법

End-to-End(검출+인식) 평가 방법

- 순차(Cascade) 평가 처리 : 검출 평가(IoU) → 인식 평가(WEM, 1-NED)

GT
Pred



IoU > 0.5

$$\text{Recall} = \frac{\text{맞춘 단어 수}}{\text{정답 단어 수}}$$

$$= \frac{1}{2} = 0.5$$

$$\text{Precision} = \frac{\text{맞춘 단어 수}}{\text{예측 단어 수}}$$

$$= \frac{1}{2} = 0.5$$

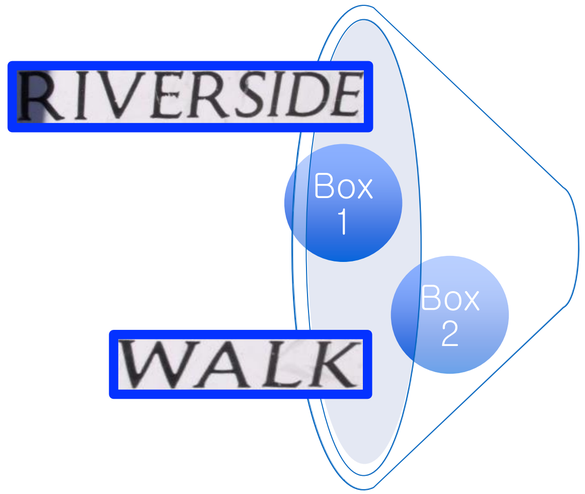
$$H_{mean} = \frac{2 \times 0.5 \times 0.5}{0.5+0.5}$$

$$= 0.5$$

3.2 기존 평가 방법

End-to-End(검출+인식) 평가 방법

- 순차(Cascade) 평가 처리 : 검출 평가(IoU) → 인식 평가(WEM, 1-NED)



IoU > 0.5

$$1-\mathrm{NED} = \frac{1}{2}\left\{(1-\frac{1}{9})+(1-\frac{0}{4})\right.$$

$$= 0.95$$

4.기존 방법의 문제점

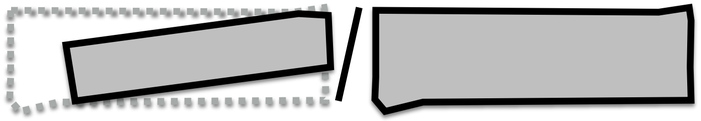
4.1 기존 방법의 문제점

정교한 성능 측정 불가 (부정확성)

- False Positive

— GT
— Pred

$$IoU = \frac{R_{Pred} \cap R_{GT}}{R_{Pred} \cup R_{GT}}$$

$$= \text{  }$$

$$= 0.54 \geq \mathbf{0.5}$$

EAST: An Efficient and Accurate Scene Text Detector (Xinyu Zhou, et. al, CVPR2017)

4.1 기존 방법의 문제점

정교한 성능 측정 불가 (부정확성)

- False Negative

GT

Pred

$$IoU = 0.39 \leq \mathbf{0.5}$$

이미지 출처: {https://rrc.cvc.uab.es/?ch=14&com=evaluation&view=method_sample&task=1&m=54589&gtv=10&file=1&eval=1&sample=304}

4.1 기존 방법의 문제점

One-to-Many 문제

- 하나의 정답 박스가 여러개의 박스로 나뉘어 예측되는 경우 (Split 이라고도 함)

4.1 기존 방법의 문제점

Many-to-One 문제

- 여러개의 정답 박스가 한개의 박스로 합쳐져 예측되는 경우 (Merge 이라고도 함)

GT

Pred

False Negative

$(\mathbf{IoU} = \mathbf{0.39} \leq \mathbf{0.5})$

4.1 기존 방법의 문제점

Many-to-One 문제

- 여러개의 정답 박스가 한개의 박스로 합쳐져 예측되는 경우 (Merge 이라고도 함)

GT

Pred

$$IoU = 0.41 \leq \mathbf{0.5}$$

False Negative

이미지 출처: {https://rrc.cvc.uab.es/?ch=15&com=evaluation&view=method_sample&task=4&m=57493&gtv=10&file=1&eval=1&sample=98}

4.1 기존 방법의 문제점

True Positive

False Negative

어느 쪽이 글자 검출을 잘 했나요?

EAST

R(1.00), P(1.00), H(1.00)

PixelLink

R(1.00), P(0.67), H(0.80)

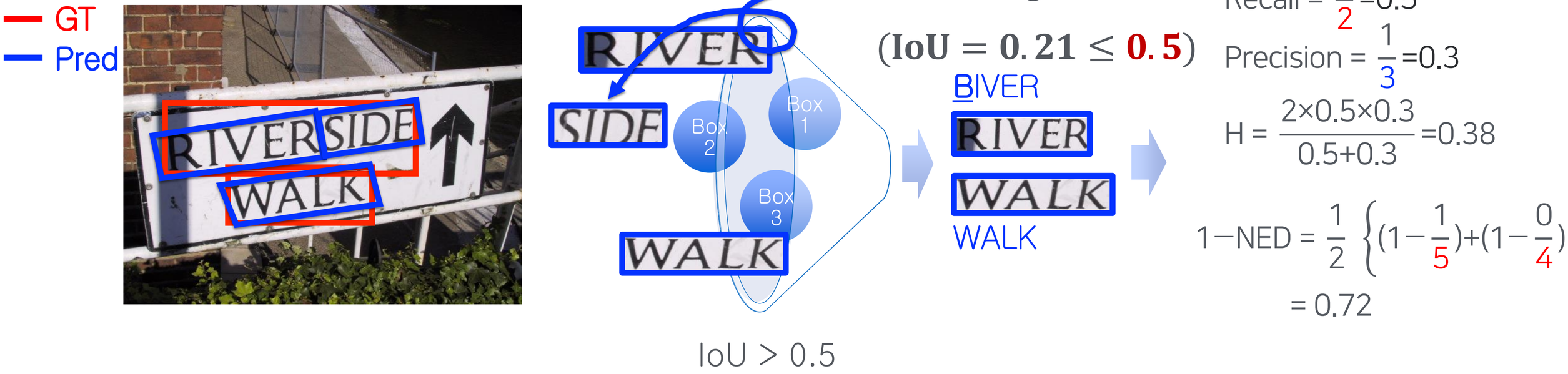
EAST: An Efficient and Accurate Scene Text Detector (Xinyu Zhou, et. al, CVPR2017)

PixelLink: Detecting Scene Text via Instance Segmentation (Dan Deng, et. al. AAAI2018)

4.2 기존 방법의 문제점

End-to-End(검출+인식) 평가 방법의 문제점

- 순차(Cascade) 평가 처리 방식으로 오류가 전파됨



4.2 기존 방법의 문제점 요약

정교한 성능 측정 불가 (부정확성)

One-to-Many, Many-to-One 대응 불가

- (객체 검출과 달리) 문자 인식에서는 자주 발생

순차(Cascade) 방식의 End-to-End 평가방법 (오류가 전파 됨)

- 검출 평가를 통과하지 못하면 인식 평가 자체가 진행되지 않음

5. 신규 방법

“PopEval”

5.1 고려 사항 (신규 평가 방법의 원칙)

End-to-End (검출+인식) 평가

- (사람처럼) 검출과 인식을 동시에 평가해야 함 (실제 서비스에 유용해야 함)

One-to-Many, Many-to-One 대응 가능해야 함

- (객체 검출과 달리) 문자 인식에서는 자주 발생

정교하게 그리고 세부적으로 성능 측정이 가능해야 함

- 개별 글자 단위로 평가

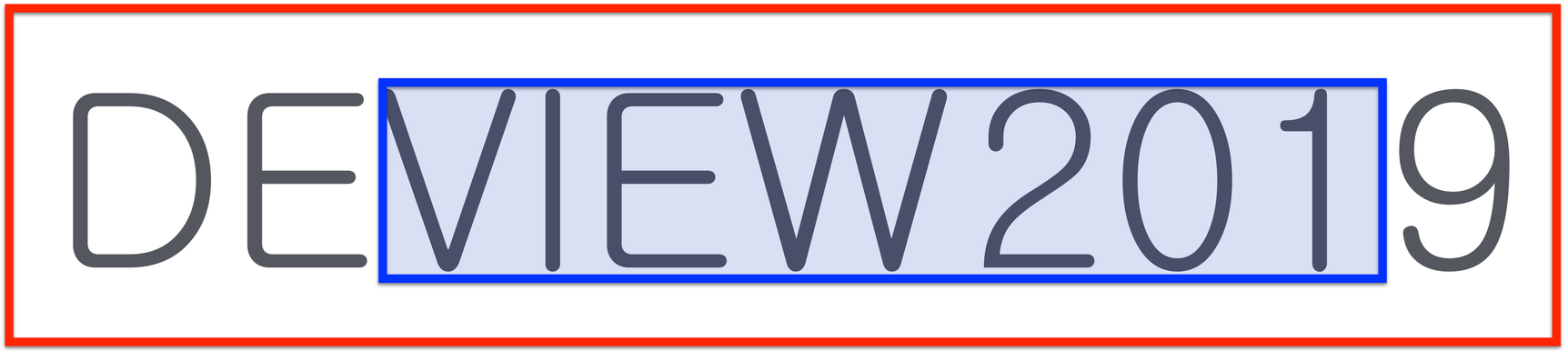
기존 평가셋과 호환되어야 함

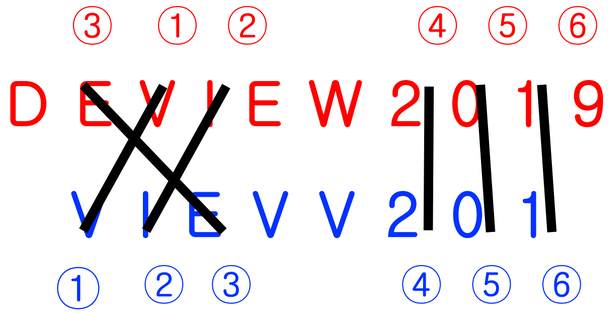
- 단어 단위의 평가셋에서도 잘 동작해야 함
- (글자 단위의 평가셋은 거의 없으며, 새로 만들기 어려움)

5.2 신규 방법

매우 간단하지만, 정확하고 정교한 방법

- 겹치는 영역의 글자 중에서, 같은 글자(=맞춘 글자)를 하나씩 지워가요!





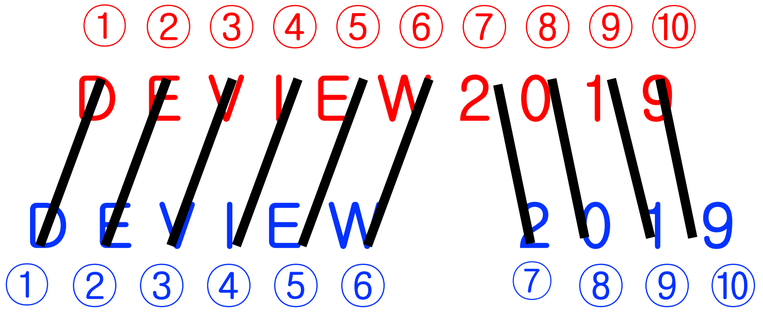
$$\text{Recall} = \frac{\text{맞춘 글자 수}}{\text{정답 글자수}} = \frac{6}{\text{len("DEVIEW2019")}} = \frac{6}{10} = 0.60$$

$$\text{Precision} = \frac{\text{맞춘 글자 수}}{\text{예측한 글자수}} = \frac{6}{\text{len("VIEVV201")}} = \frac{6}{8} = 0.75$$

5.2 신규 방법

One-to-Many 문제도 처리 가능

- 겹치는 영역의 글자 중에서, 같은 글자(=맞춘 글자)를 하나씩 지워가요!



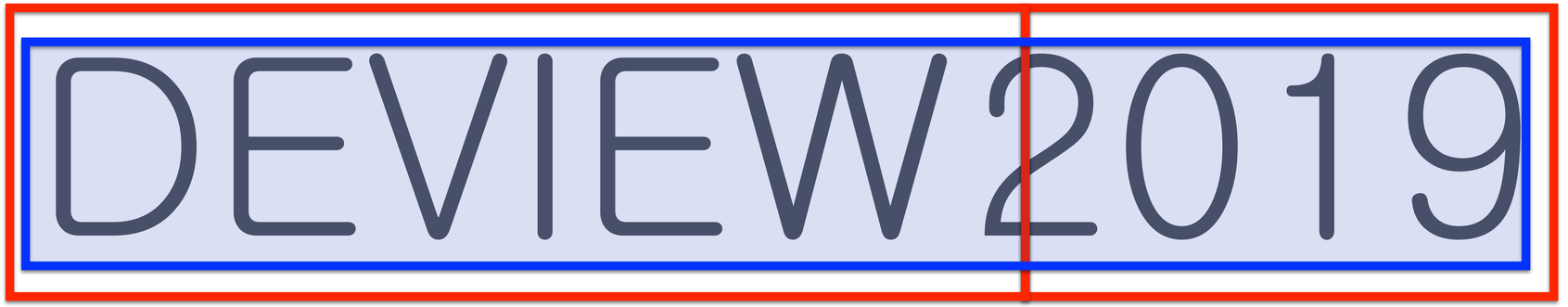
$$\text{Recall} = \frac{\text{맞춘 글자 수}}{\text{정답 글자수}} = \frac{10}{\text{len(“DEVIEW2019”)}} = \frac{10}{10} = 1.00$$

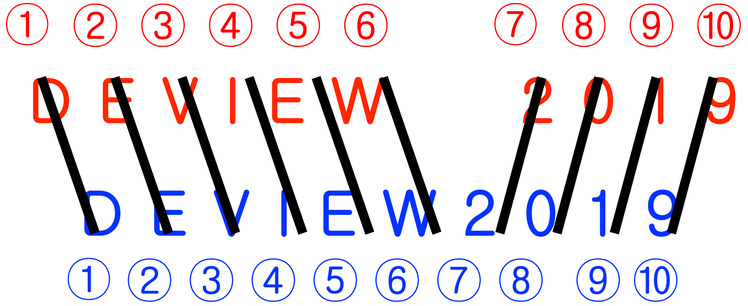
$$\text{Precision} = \frac{\text{맞춘 글자 수}}{\text{예측한 글자수}} = \frac{10}{\text{len(“DEVIEW2019”)}} = \frac{10}{10} = 1.00$$

5.2 신규 방법

Many-to-One 문제도 처리 가능

- 겹치는 영역의 글자 중에서, 같은 글자(=맞춘 글자)를 하나씩 지워가요!





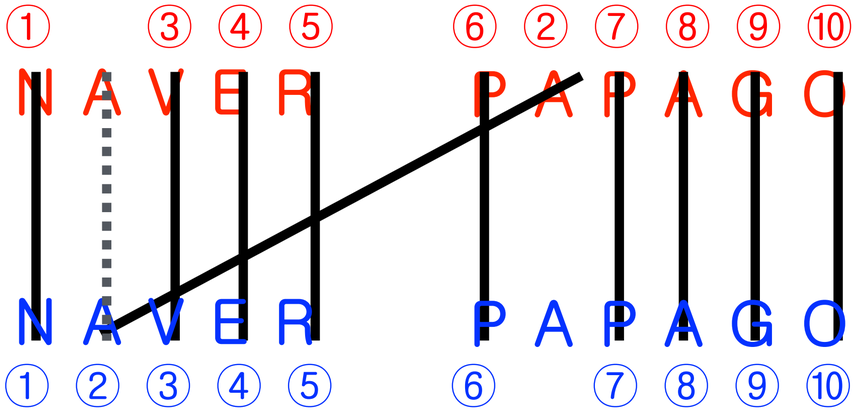
$$\text{Recall} = \frac{\text{맞춘 글자 수}}{\text{정답 글자수}} = \frac{10}{\text{len(“DEVIEW2019”)}} = \frac{10}{10} = 1.00$$

$$\text{Precision} = \frac{\text{맞춘 글자 수}}{\text{예측한 글자수}} = \frac{10}{\text{len(“DEVIEW2019”)}} = \frac{10}{10} = 1.00$$

6.3 신규 방법

엣지 케이스

- 제거해야 할 글자가 중복될 경우 어떤 글자부터 제거할 것인가? (모호함)



$$\text{Recall} = \frac{\text{맞춘 글자 수}}{\text{정답 글자수}} = \frac{10}{\text{len(“NAVERPAPAGO”)}} = \frac{10}{11} = 0.91$$

$$\text{Precision} = \frac{\text{맞춘 글자 수}}{\text{예측한 글자수}} = \frac{10}{\text{len(“NAVERPAPAGO”)}} = \frac{10}{11} = 0.91$$

6.3 신규 방법

DEVIEW 2019

엣지 케이스

- 제거해야 할 글자가 중복될 경우 어떤 글자부터 제거할 것인가? (모호함)

1. (논란의 여지가 없는) 중복이 없는 글자 박스부터 먼저 제거
2. 정답과의 교집합 영역이 클수록 먼저 제거

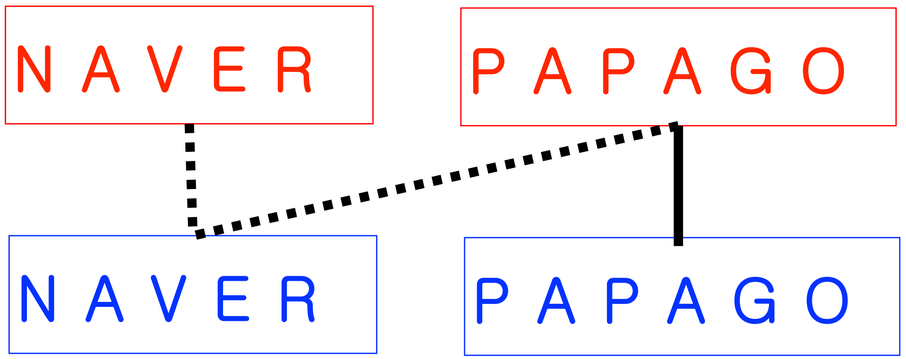
6.3 신규 방법

DEVIEW 2019

엣지 케이스

- 제거해야 할 글자가 중복될 경우 어떤 글자부터 제거할 것인가? (모호함)

1. 중복이 없는 박스 우선
2. 교집합이 클수록 우선

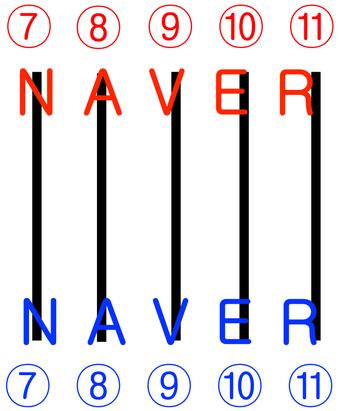


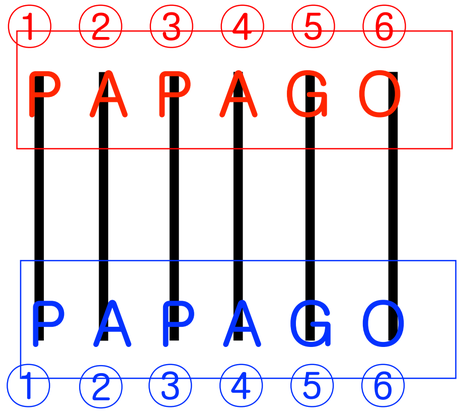
6.3 신규 방법

엣지 케이스

- 제거해야 할 글자가 중복될 경우 어떤 글자부터 제거할 것인가? (모호함)

1. 중복이 없는 박스 우선
2. 교집합이 클수록 우선





$$\text{Recall} = \frac{\text{맞춘 글자 수}}{\text{정답 글자수}} = \frac{11}{\text{len(“NAVERPAPAGO”)}} = \frac{11}{11} = 1.00$$

$$\text{Precision} = \frac{\text{맞춘 글자 수}}{\text{예측한 글자수}} = \frac{11}{\text{len(“NAVERPAPAGO”)}} = \frac{11}{11} = 1.00$$

5.3 기존 방법과 비교 (End-to-End)

DEVIEW 2019

— GT

— Pred

기존(H, 1-NED : 0.0)

신규제안(PopEval : 0.91)

5.3 기존 방법과 비교 (End-to-End)

— GT
— Pred

기존(H : 0.0, 1-NED : 0.33)

신규제안(PopEval : 1.00)

5.3 기존 방법과 비교 (End-to-End)

GT

Pred

기존(H, 1-NED : 0.0)

신규제안(PopEval : 0.91)

DEVIEW
2019

6. 신규 방법에 대한 검증 실험
(믿을만 한가?)

6.1 신규 방법에 대한 검증 사항

DEVIEW 2019

One-to-Many, Many-to-One 문제는 얼마나 발생하나?

- 전체 박스의 2~9%에 해당하며,
- 이는 리더보드 Top10 내 순위에 영향을 주는 유의미한 수치임

기존 평가셋(=단어 단위)과 호환 가능한가?

- 평가 방식은 글자 단위이지만, 단어 단위의 평가셋과 호환이 가능

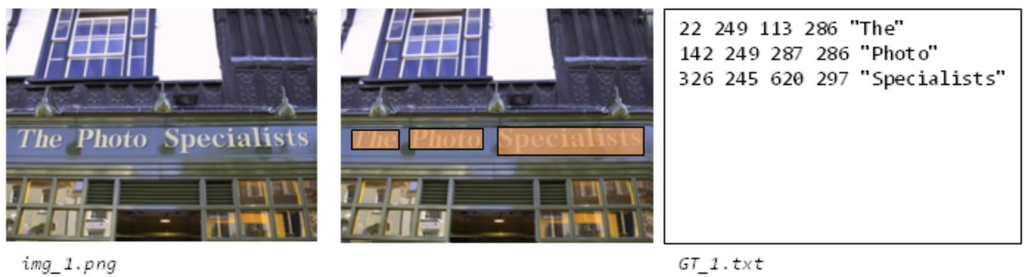
신규 평가 방법은 믿을만 한가? (신규 평가 방법에 대한 평가)

- 피어슨 통계 분석 결과, 기존 방법보다 우수함

6.2 검증을 위한 실험 환경

DEVIEW 2019

실험 데이터 (IC13, IC15 평가셋)



6.2 검증을 위한 실험 환경

DEVIEW 2019

예측 모델 : 검출기(EAST, PixelLink) x 인식기(GRCNN, ASTER)

EAST: An Efficient and Accurate Scene Text Detector

Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang

Megvii Technology Inc., Beijing, China
{zxy, yaocong, wenhe, wangyuzhi, zsc, hwr, liangjiajun}@megvii.com

PixelLink: Detecting Scene Text via Instance Segmentation

Dan Deng[1,3*], **Haifeng Liu**[1], **Xuelong Li**[4], **Deng Cai**[1,2]
[1]State Key Lab of CAD&CG, College of Computer Science, Zhejiang University
[2]Alibaba-Zhejiang University Joint Institute of Frontier Technologies
[3]CVTE Research
[4]Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences
dengdan.zju@gmail.com {haifengliu,dcai}@zju.edu.cn xuelong_li@opt.ac.cn

EAST: An Efficient and Accurate Scene Text Detector (Xinyu Zhou, et. al, CVPR2017)

PixelLink: Detecting Scene Text via Instance Segmentation (Dan Deng, et. al. AAAI2018)

6.2 검증을 위한 실험 환경

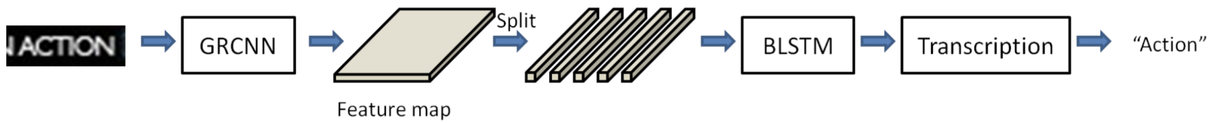
DEVIEW 2019

예측 모델 : 검출기(EAST, PixelLink) x 인식기(GRCNN, ASTER)

Gated Recurrent Convolution Neural Network for OCR

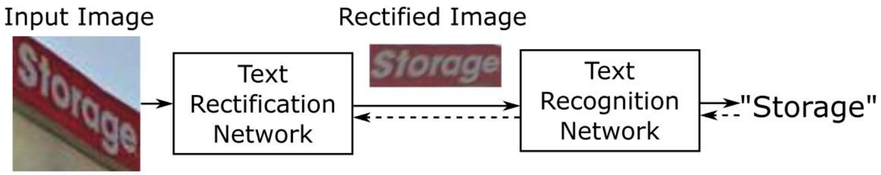
Jianfeng Wang*
Beijing University of Posts and Telecommunications
Beijing 100876, China
jianfengwang1991@gmail.com

Xiaolin Hu
Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Computer Science and Technology
Center for Brain-Inspired Computing Research (CBICR)
Tsinghua University, Beijing 100084, China
xlhu@tsinghua.edu.cn



ASTER: An Attentional Scene Text Recognizer with Flexible Rectification

Baoguang Shi, Mingkun Yang, Xinggang Wang, Pengyuan Lyu, Cong Yao, and Xiang Bai



Gated Recurrent Convolution Neural Network for OCR (Wang, et. al, NIPS2017)

ASTER: An Attentional Scene Text Recognizer with Flexible Rectification (Baoguang Shi, et. al. PAMI2019)

6.3 신규 방법에 대한 검증 실험

One-to-Many, Many-to-One 문제는 얼마나 발생하나?

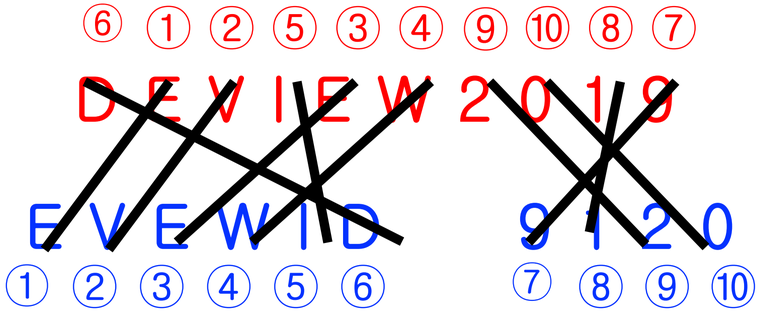
- 전체 박스의 2~9%에 해당하며,
- 이는 리더보드 Top10 내 순위에 영향을 주는 유의미한 수치임

	Split Detections (one-to-many)	Merged GTs (many-to-one)
EAST - ICDAR2013	3.84%	1.46%
PIXEL - ICDAR2013	6.09%	3.29%
EAST - ICDAR2015	1.13%	1.54%
PIXEL - ICDAR2015	2.05%	0.35%

6.3 신규 방법에 대한 검증 실험

기존 평가셋(=단어 단위)과 호환이 가능한가?

- 단어 단위에서 개별 글자의 순서가 뒤바뀌는 문제점 (Permutation)



$$\text{Recall} = \frac{\text{맞춘 글자 수}}{\text{정답 글자수}} = \frac{10}{\text{len(“DEVIEW2019”)}} = \frac{10}{10} = 1.00$$

$$\text{Precision} = \frac{\text{맞춘 글자 수}}{\text{예측한 글자수}} = \frac{10}{\text{len(“DEVIEW2019”)}} = \frac{10}{10} = 1.00$$

6.3 신규 방법에 대한 검증 실험

DEVIEW 2019

기존 평가셋(=단어 단위)과 호환이 가능한가?

- 단어 단위에서 개별 글자의 순서가 뒤바뀌는 경우는 거의 없음

AMONG THE RECOGNITION RESULTS WHICH COMPOSED OF THE SAME ALPHANUMERIC COMPONENTS AS GT, THE PROPORTION THAT DOES NOT EXACTLY MATCH GT.

	ICDAR2013	ICDAR2015
ASTER	0.00%	0.05%
GRCNN	0.00%	0.14%

6.3 신규 방법에 대한 검증 실험

기존 평가셋(=단어 단위)과 호환이 가능한가?

- 단어 단위에서 개별 글자의 순서가 뒤바뀌는 경우의 예

— GT
— Pred

6.3 신규 방법에 대한 검증 실험

DEVIEW 2019

기존 평가셋(=단어 단위)과 호환이 가능한가?

- 단어 단위와 글자 단위의 평가 결과 비교

6.3 신규 방법에 대한 검증 실험

기존 평가셋(=단어 단위)과 호환이 가능한가?

- 단어 단위와 글자 단위의 평가 결과 차이 없음 (호환 가능한 수준)

For ICDAR2013 Test Dataset

	Word Level	Character Level	Diff
EAST - ASTER	0.8649	0.8616	0.0033
PIXEL - GRCNN	0.8562	0.8531	0.0031
EAST - ASTER	0.8540	0.8513	0.0027
PIXEL - GRCNN	0.8552	0.8538	0.0014

For ICDAR2015 Test Dataset

	Word Level	Character Level	Diff
EAST - ASTER	0.8017	0.7991	0.0026
PIXEL - GRCNN	0.7696	0.7661	0.0035
EAST - ASTER	0.7792	0.7783	0.0009
PIXEL - GRCNN	0.8003	0.7986	0.0017

6.3 신규 방법에 대한 검증 실험

DEVIEW
2019

신규 평가 방법은 믿을만 한가? (신규 평가 방법에 대한 평가)

- 3명의 평가자, 기존 방법과 신규 방법에 대한 점수 부여 (5점 척도)
- 글자 박스 개수 (IC13 : 1,015개, IC15 : 2,077개)

6.3 신규 방법에 대한 검증 실험

신규 평가 방법은 믿을만 한가? (신규 평가 방법에 대한 평가)

- 피어슨 통계 분석 결과, 기존 방법보다 우수함

For ICDAR2013 Test Dataset

	Vocab	AP	1-NED	PopEval at word	PopEval at character
EAST - ASTER	0.7858	0.4595	0.8884	0.9305	0.9340
EAST - GRCNN	0.7910	0.4437	0.8800	0.9457	0.9461

For ICDAR2015 Test Dataset

	Vocab	AP	1-NED	PopEval at word	PopEval at character
EAST - ASTER	0.7776	0.5792	0.8124	0.9272	0.9213
EAST - GRCNN	0.6870	0.5410	0.7262	0.8221	0.8204

6.4 신규 방법에 대한 요약 및 정리

DEVIEW 2019

정확하고 정교한 End-to-End(검출+인식) 평가 방법

- 사람 평가방식과 가장 유사함
- 글자 단위로 평가하기 때문에 정교함 (One-to-Many, Many-to-One 대응가능)
- 최신 연구(Arbitrary text shape) 대응 가능

기존 평가셋(단어 단위)와 호환 가능

현업에 바로 사용 가능

- 논문 : https://arxiv.org/abs/1908.11060
- 코드 : https://github.com/naver/popeval

DEVIEW
2019

Q & A

DEVIEW
2019

Thank You

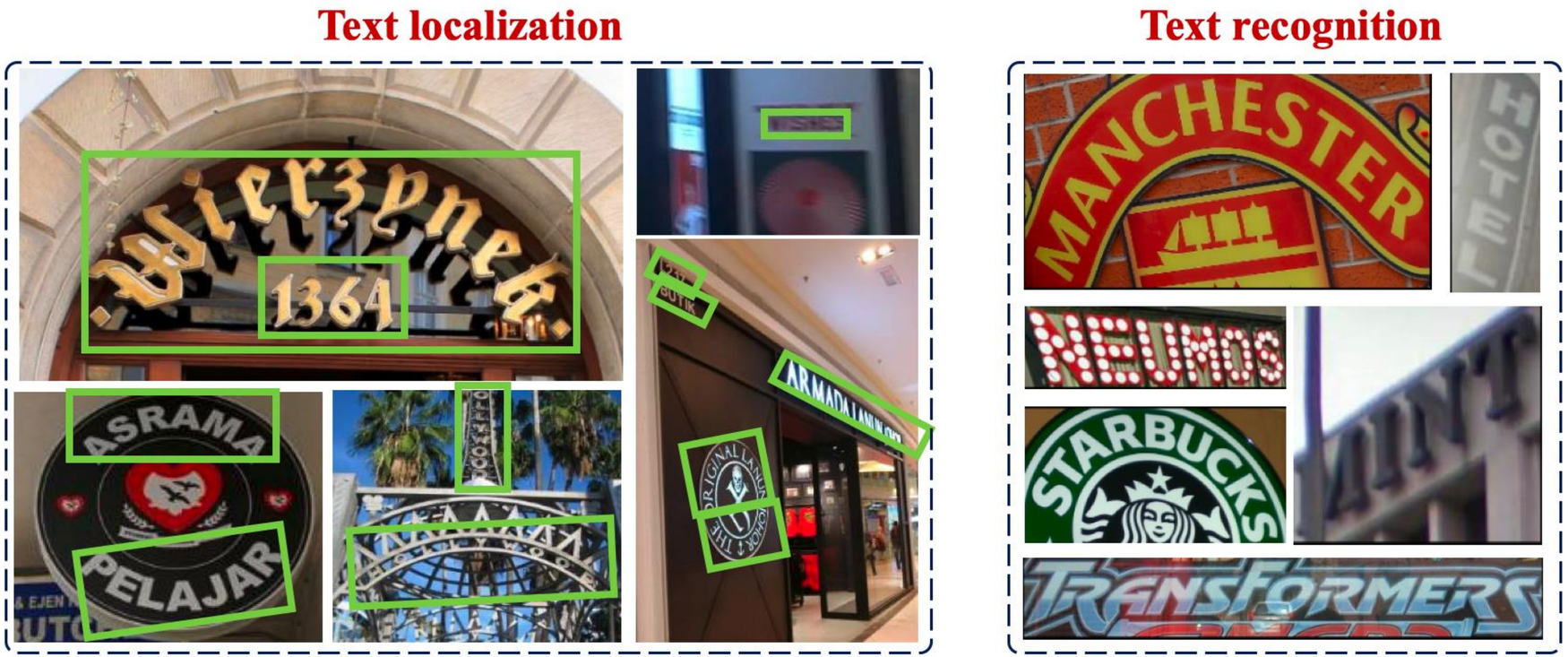
DEVIEW
2019

Appendix

2.3 문자인식 연구동향

DEVIEW 2019

최신 연구(CVPR2019)의 대부분이 Arbitrary Text Shape에 관한 것



Xiang Bai et al. ICDAR2019

3.1 사전지식 (Recall, Precision)

암 진단에 대한 평가

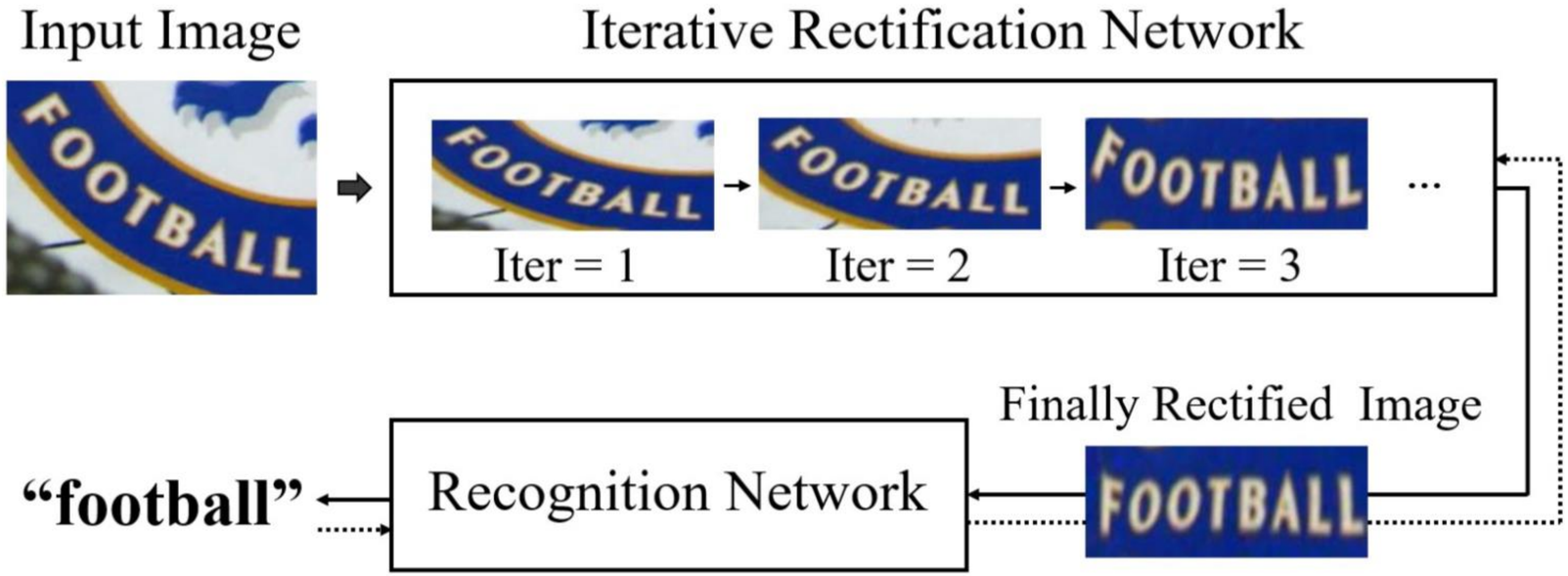
이미지 출처: {https://www.ownyourhealthwa.org/wp-content/uploads/HowToShareYourOwnResearchWithYourDoctor-1024x943.jpg}

4.1 기존 방법의 문제점

DEVIEW 2019

정교한 성능 측정 불가 (부정확성)

- 최근에는 Curved text를 인식할 수 있음



ESIR: End-to-end Scene Text Recognition via Iterative Image Rectification (Fangneng Z. et. al. CVPR2019)

4.1 기존 방법의 문제점

DEVIEW 2019

정교한 성능 측정 불가 (부정확성)

- 최근에는 Curved text를 인식할 수 있음

2D Attentional Irregular Scene Text Recognizer (Pengyuan L, et. al. arXiv 2019)

6.3 신규 방법에 대한 검증 실험

기존 평가셋(=단어 단위)과 호환이 가능한가?

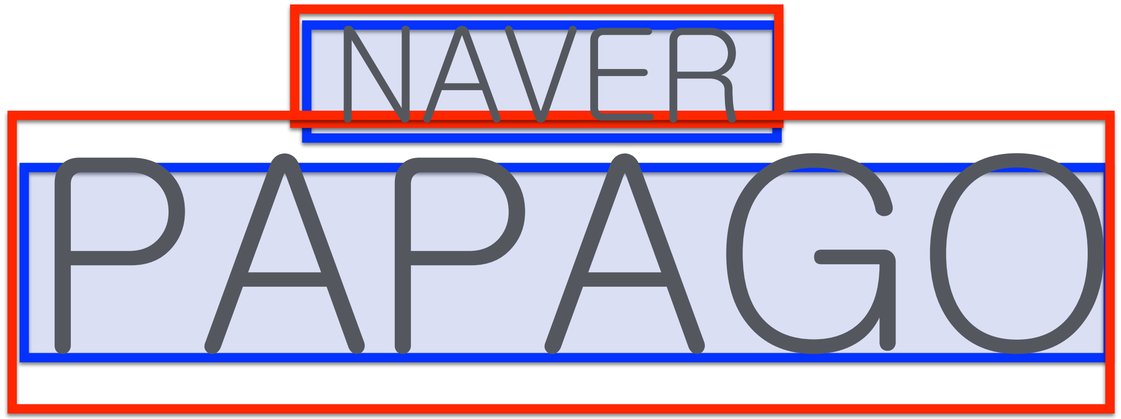
- 단어 단위에서 개별 글자의 순서가 뒤바뀌는 경우의 예

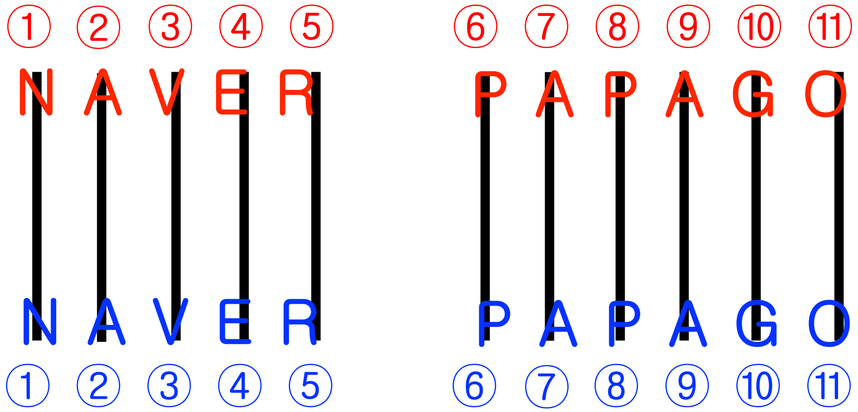
— GT
— Pred

6.3 신규 방법

엣지 케이스

- 제거해야 할 글자가 중복될 경우 어떤 글자부터 제거할 것인가? (모호함)





$$\text{Recall} = \frac{\text{맞춘 글자 수}}{\text{정답 글자수}} = \frac{11}{\text{len(“NAVERPAPAGO”)}} = \frac{11}{11} = 1.00$$

$$\text{Precision} = \frac{\text{맞춘 글자 수}}{\text{예측한 글자수}} = \frac{11}{\text{len(“NAVERPAPAGO”)}} = \frac{11}{11} = 1.00$$

예상 질문

DEVIEW
2019

IoU를 낮추고, 1-NED로 측정