

Optical Character Recognition

01

INDEX



01

Introduction

- About OCR
- Advanced Work
- Project objective

02

About Model

- Detector
- Super Resolution
- Recognizer
- Post-Processing

03

Preview Demo Model

- Structure
- Demonstration

04

Difficulties

- Detection
- Super Resolution
- Recognition
- Post-Processing

05

Conclusion

- Future Directions
- Q&A

02

+

01

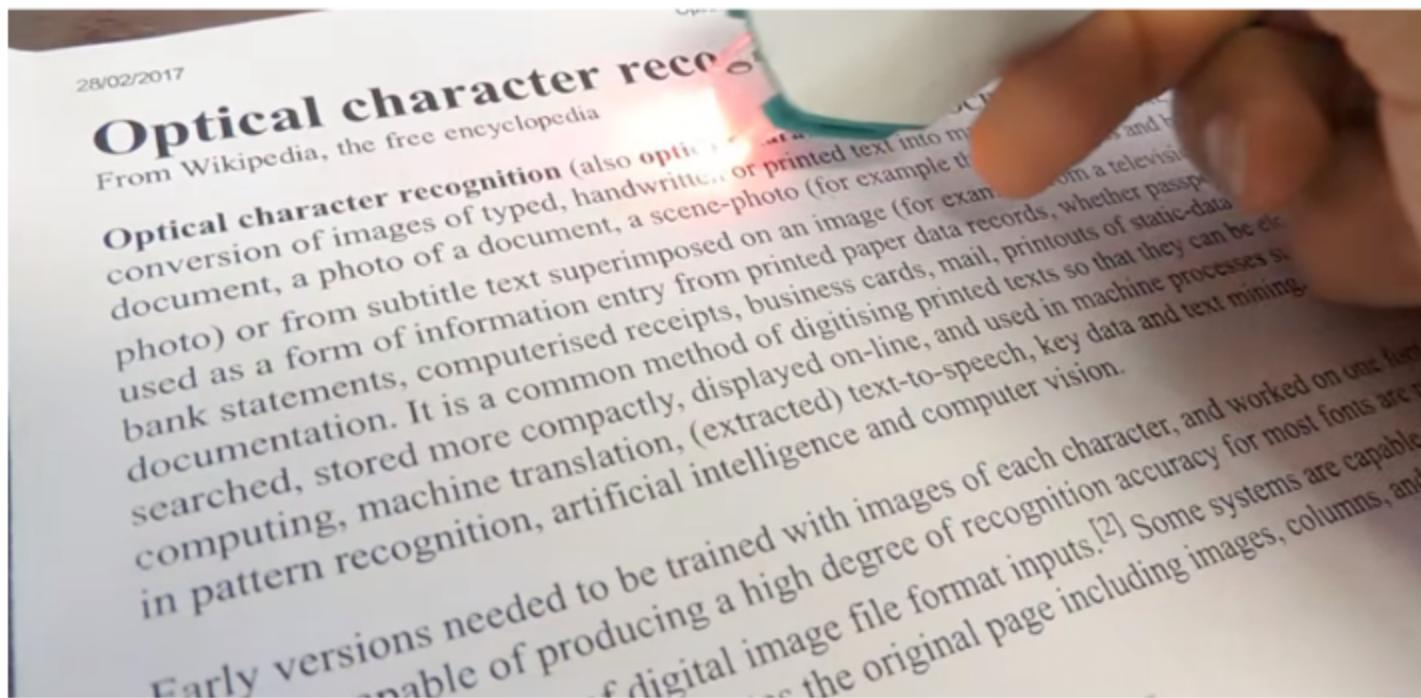
Introduction

- OCR
- Advanced Work
- Project objective

03

01. Introduction

- OCR (광학 문자 인식)



Display Image

경상남도		
우 641-702	경남 창원시 사용동 1번지	/전화 055-211-4835 /전송 055-211-4819
문화예술과	과장 이재룡	사무관 하재봉 담당자 망기현
문서번호 문예86700-10052		
시행일자 2003.01.09	(3)년	시장
공개여부 ()		부시장
발 름 경상남도 김해시 가야권종합개발	16:54	제 소장
한 조	번호 5037	과장
	처리과 문화예술과	전화
	담당자 망기현	fax
제 목 김해 대성동고분군 경비사업 결과변경 승인	심사자	결정
1. 문정86700-5434(11.20)호 및 사적86743-1961(12.23)호와 관련입니다.		
2. 상기호와 관련하여가이역사문화환경경비사업으로 추진중인 김해대성동 고분군(사적 제341호)경비사업 설계변경을 위 시 변경연과 같이 승인하니 사업추 진에 풀점을 기하기 바리며,		
3. 다만, 사업대상부지(보도확장 구간)가 대성동고분군과 연접한 부지임에 따라 관계전문가의 입회하여 공사를 추진하여 주시고, 유구가 발견될 경우에는 관련법령에 의거 필요한 조치를 강구하시기 바랍니다. 끝.		

94

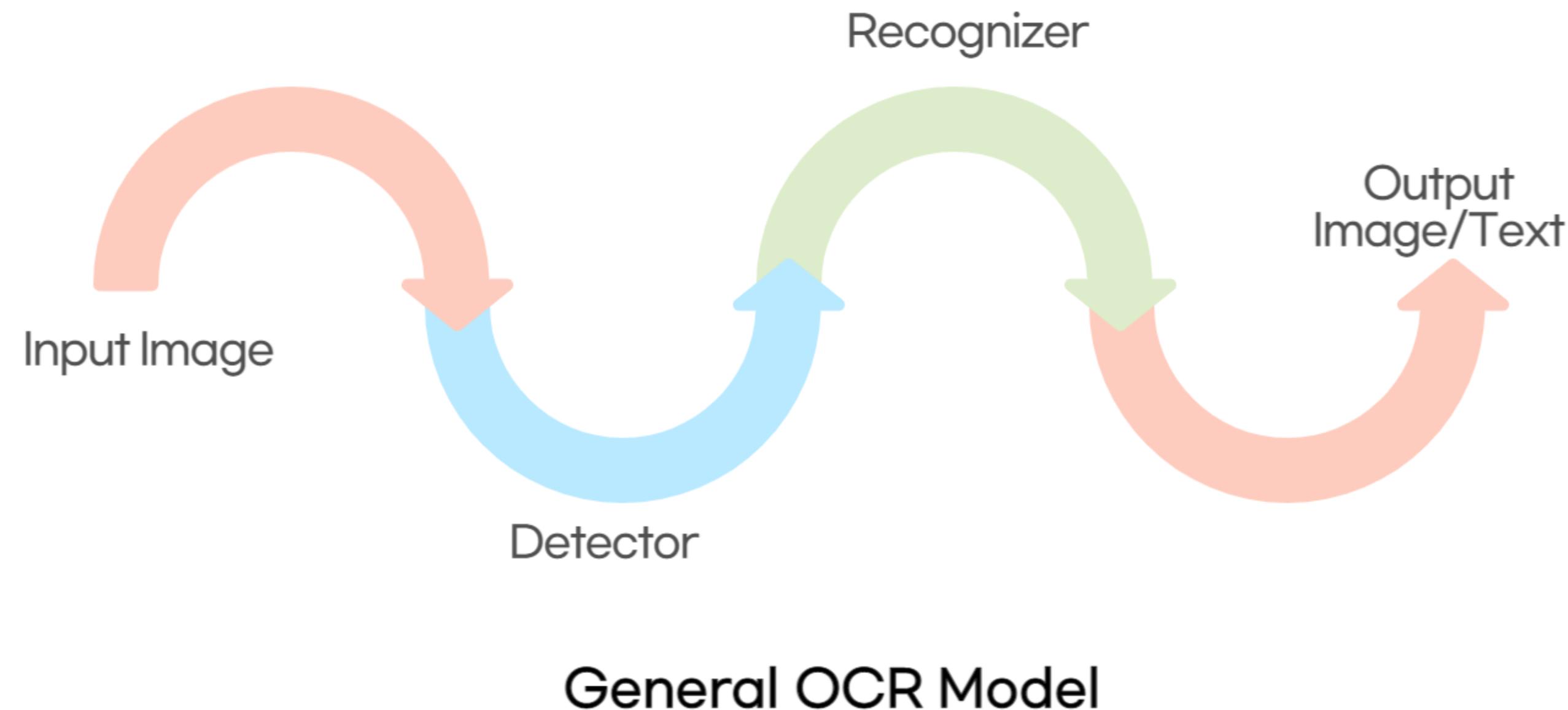
경상남도지
전결 과장 이재룡

Recognition

경상남도 우 641-702 경남 창원시 사용동 1번지 /전화 055-211-4835 /전송 055-211-4819 문화예술과 과장이재룡 사무관 하재봉 담당자 망기현 문서번호 문예86700-10052 선결 시장 지시 시행일자 2003.01.09 (3)년 공개여부 () 일자 2003.01.10 부시장 결재 시간 16:54 발 름 경상남도 김해시 가야권종합개발 소장 번호 5037 공람 과장 전결 처리과 문화정비과 참 조 김병옥 담당자 심사자 심사일 제 목김해 대성동고분군 정비사업 설계변경 승 1. 문정86700-5434(11.20)호 및 사적86743-1981112.23)호와 관련입니다. 2. 상기호와 호하여가야역사문화환경경비사업으로 추진중인 김해대성동 고분군(사적 제341호)정비사업 설계변경을 위 시 변경연과 같이 승인하니 사업추 진에 철저를 기하기 바라며, 3. 다만, 사업대상부지(보도확장 구간)가 대성동고분군과 연접한 부지임에 따라 관계전문가의 입회하여 공사를 추진하여 주시고, 유구가 발견될 경우에는 관련법령에 의거 필요한 조치를 강구하시기 바랍니다. 끝. 경상남도지 전결 과장 이재룡

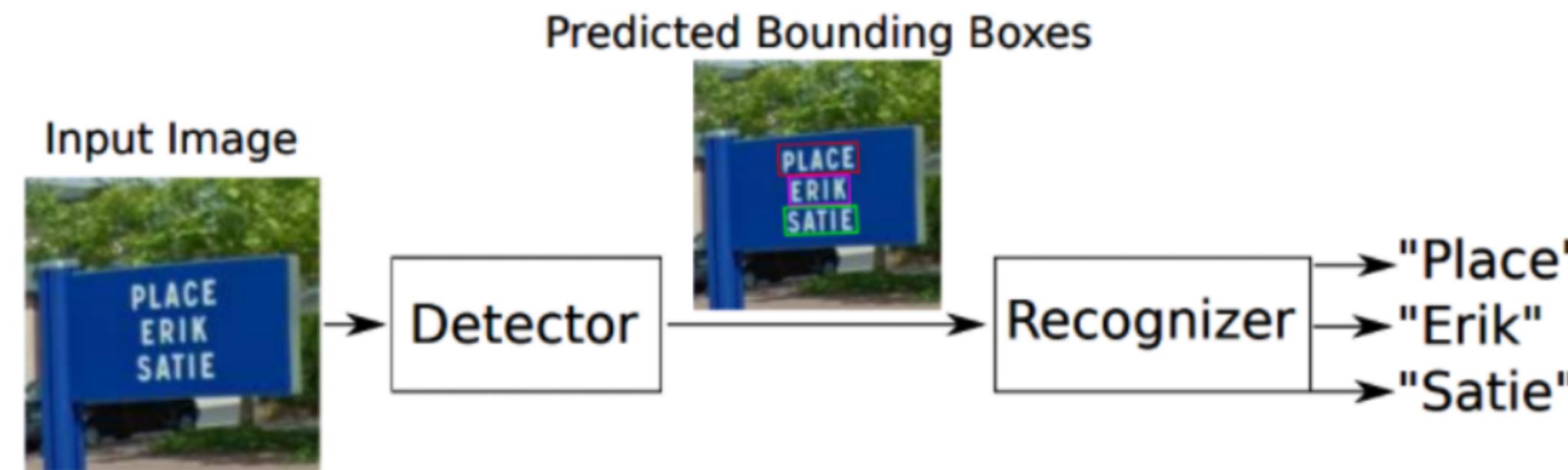
01. Introduction

- OCR (광학 문자 인식)



01. Introduction

- OCR (광학 문자 인식)



OCR = Text detection + Text recognition

06

01. Introduction

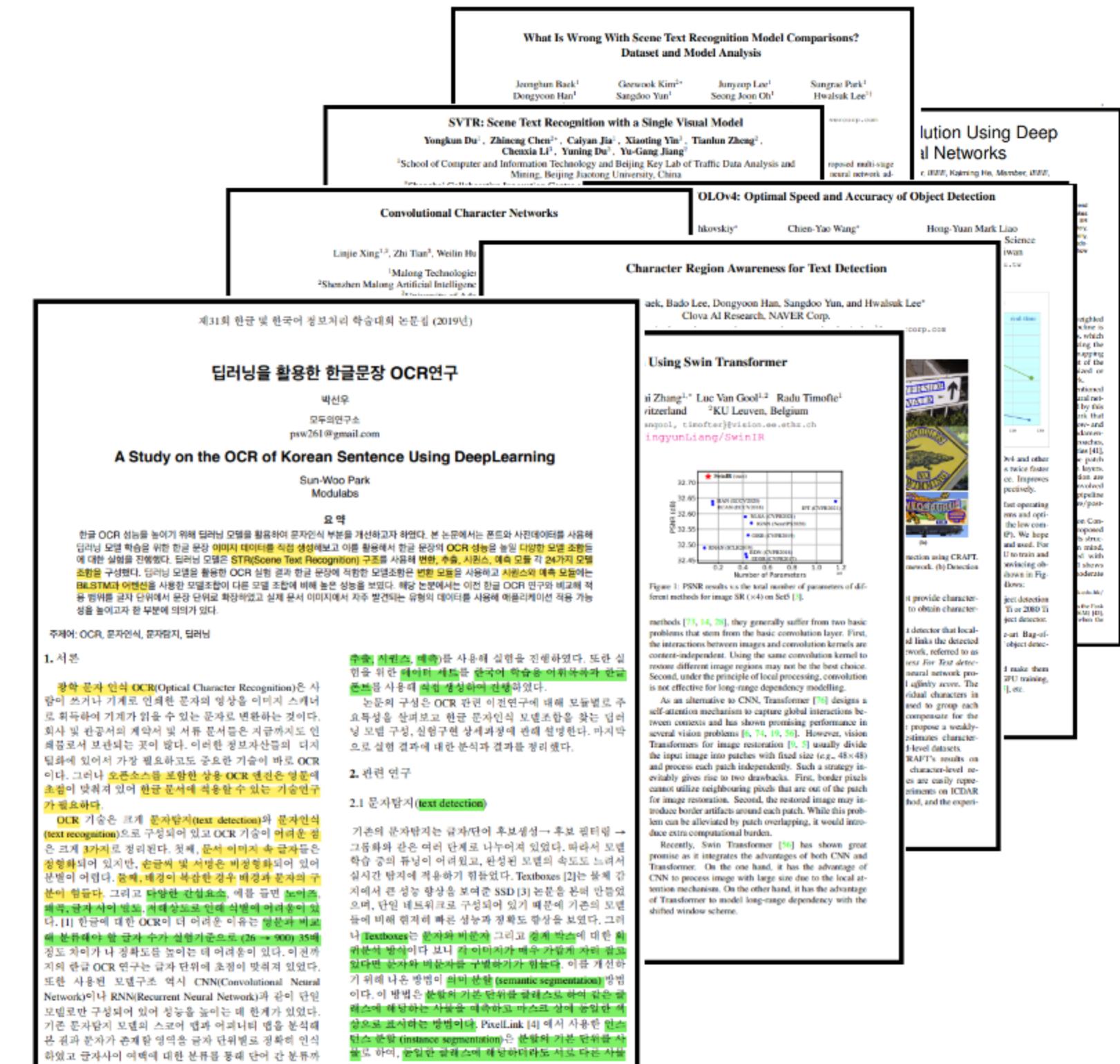
- Advanced Work

○ Background

Detector 모델 탐색 : YOLO, CRAFT, Detectron2, etc..

Recognition 모델 탐색 : STARNet, SVTR, SGTR, etc...

한글 OCR의 문제점에 대한 논문



01. Introduction

- Advanced Work

○ Approach

성능 기반 모델 선정 후 구현

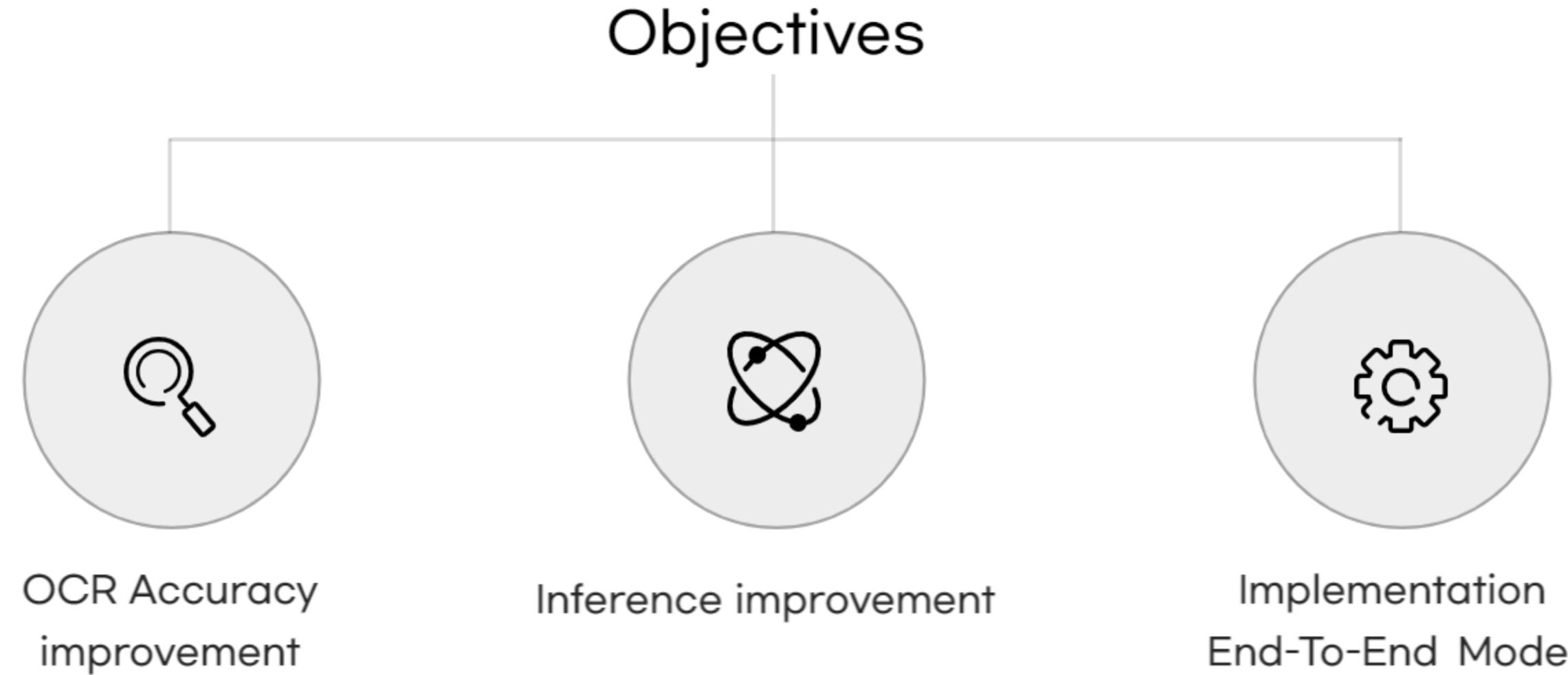
논문 관련 한글 인식률을 높이기 위한 방안 모색

Model	size (pixels)	method		English regular			English unregular			Chinese Scene	Params (M)	Speed (ms)
		IC13 SVT	III5k 3000	IC15 1811	SVTP	CUTE80	Avg_6	IC15 2077	IC13 1015			
CRNN[Shi <i>et al.</i> , 2017]		91.1	81.6	82.9	69.4	70.0	65.5	53.4	8.3	6.3	95.70	90.13
Rosetta[Borisvuk <i>et al.</i> , 2018]		90.9	84.7	84.3	71.2	73.8	69.2	-	44.3	10.5	96.28	91.02
YOLOv5n	640	28.0	45.7	45	6.3	0.6	1.9	4.5	-	40.0	95.81	91.61
YOLOv5s	640	37.4	56.8	98	6.4	0.9	7.2	16.5	-	11.2	96.74	92.24
YOLOv5m	640	45.4	64.1	224	8.2	1.7	21.2	49.0	-	50.6		
YOLOv5l	640	49.0	67.3	430	10.1	2.7	46.5	109.1	-	-		
YOLOv5x	640	50.7	68.9	766	12.1	4.8	86.7	205.7	-	-		
YOLOv5n6	1280	36.0	54.4	153	8.1	2.1	3.2	4.6	-	-		
YOLOv5s6	1280	44.8	63.7	385	8.2	3.6	12.6	16.8	-	-		
YOLOv5m6	1280	51.3	69.3	887	11.1	6.8	35.7	50.0	-	-		
YOLOv5l6	1280	53.7	71.3	1784	15.8	10.5	76.8	111.4	-	-		
YOLOv5x6 + TTA	1280	55.0	72.7	3136	26.2	19.4	140.7	209.8	-	-		
	1536	55.8	72.7	-	-	-	-	-	-	-		

ng and evaluation settings. This inconsistency is corrected by the original papers and also show our best model we have found, and, PRI denote MJSynth [11], SynthText [8], each benchmark is shown in **bold**.

01. Introduction

- Project Objectives



02

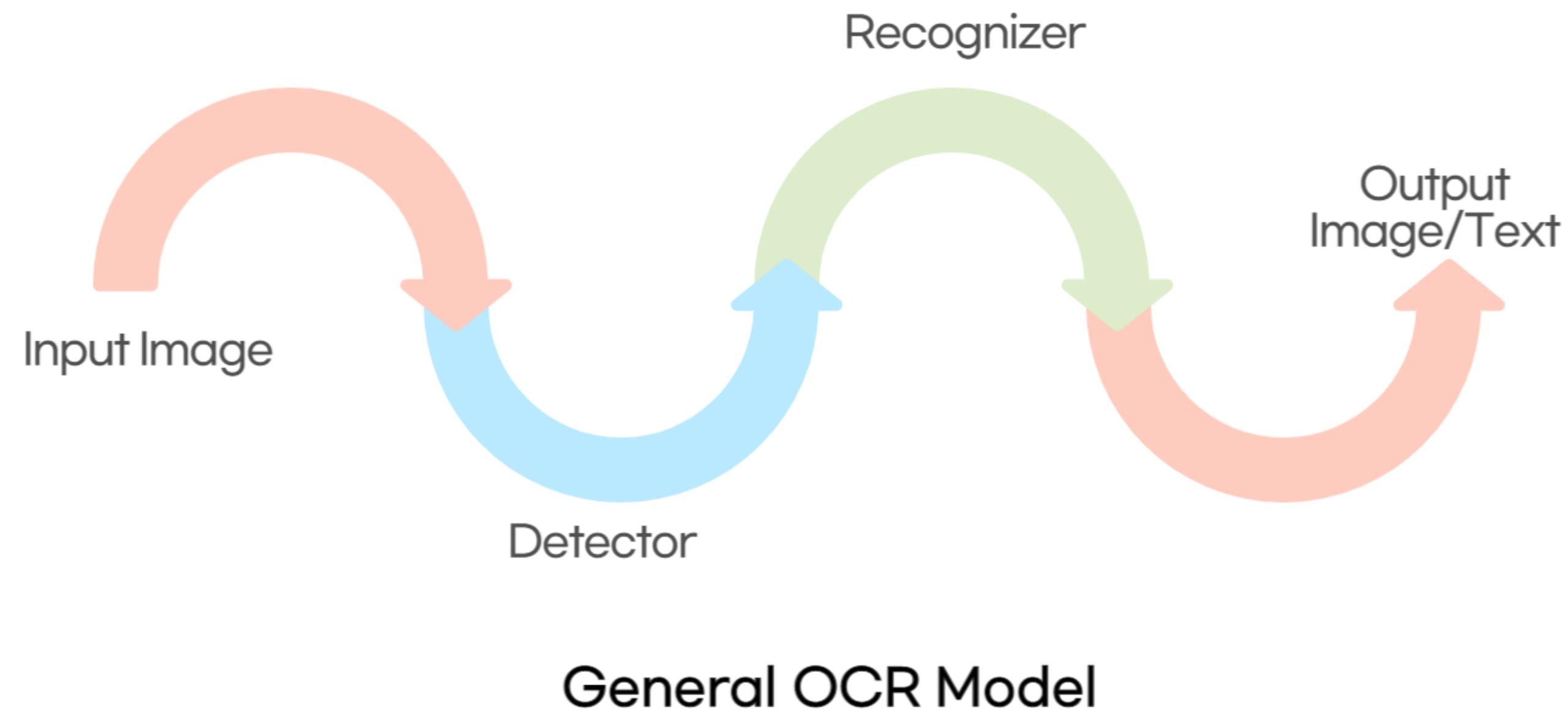
About Model

- Detector
- Super Resolution
- Recognizer
- Post-Processing

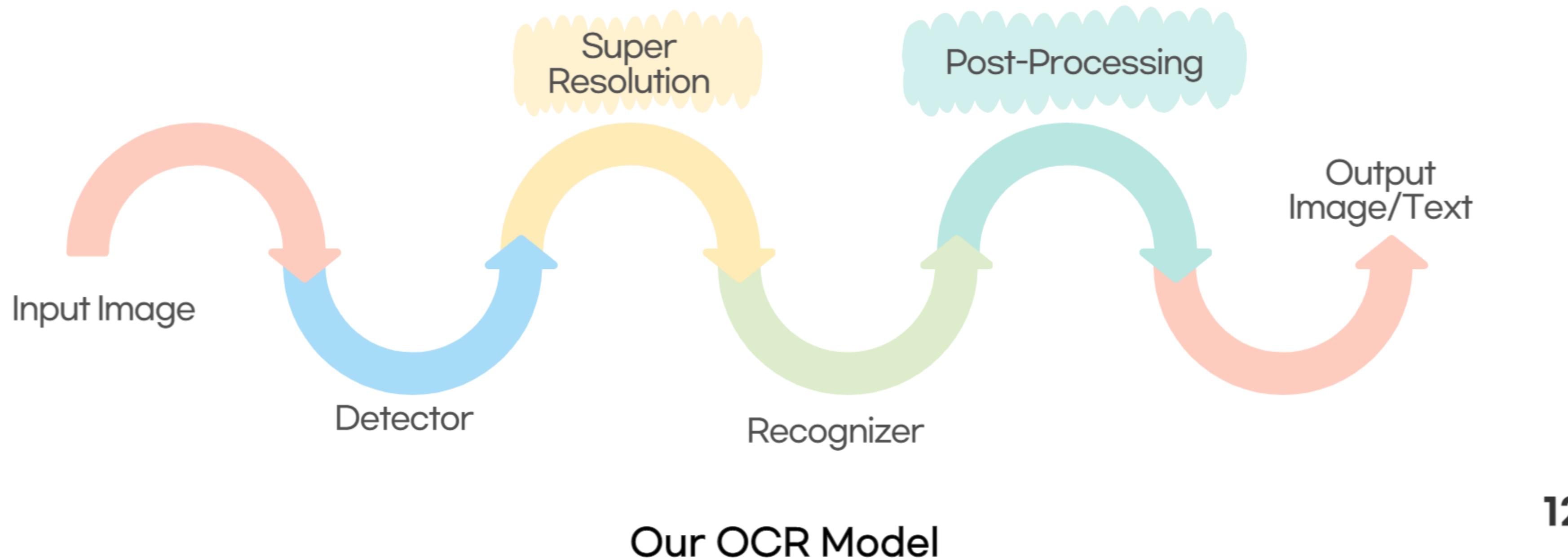
+

10

02. About Model

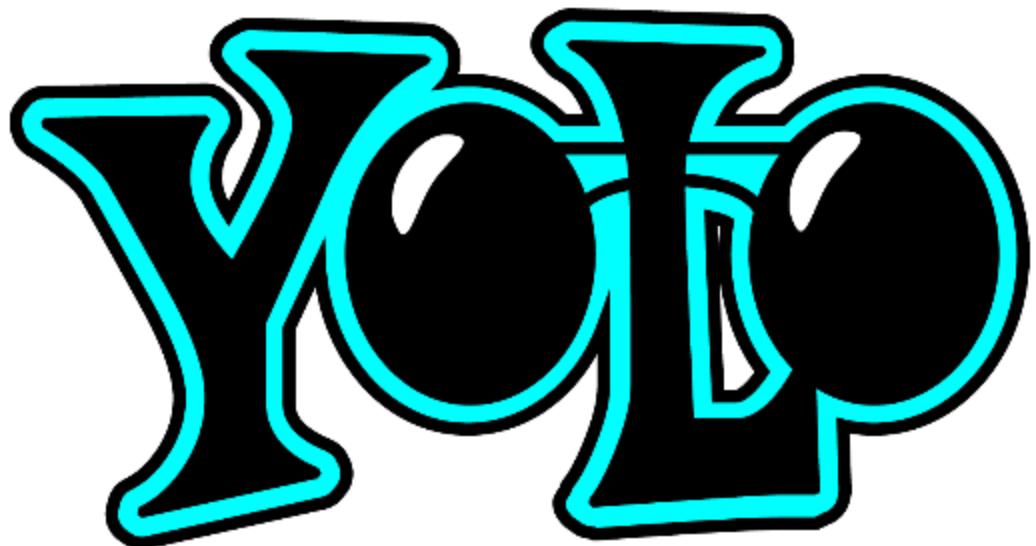


02. About Model



02. About Model

- Detector



02. About Model

- Detector

Clova

YOLO



Tesseract OCR



Detectron2

02. About Model

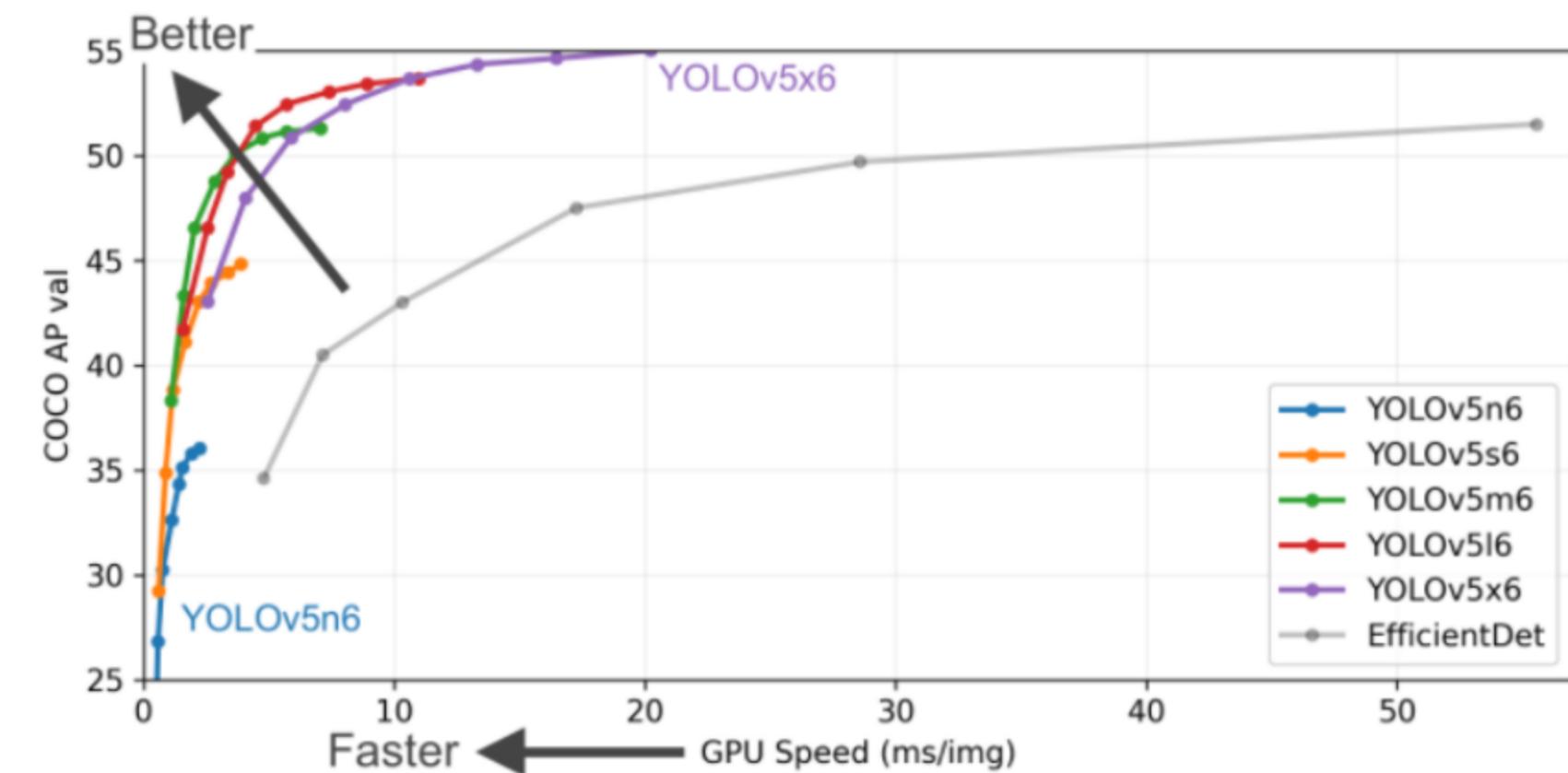
- Detector

YOLOv5 by 

02. About Model

- Detector : YOLOv5

Model	size (pixels)	mAP ^{val} 0.5:0.95	mAP ^{val} 0.5	Speed CPU b1 (ms)	Speed V100 b1 (ms)	Speed V100 b32 (ms)	params (M)	FLOPs @640 (B)
YOLOv5n	640	28.0	45.7	45	6.3	0.6	1.9	4.5
YOLOv5s	640	37.4	56.8	98	6.4	0.9	7.2	16.5
YOLOv5m	640	45.4	64.1	224	8.2	1.7	21.2	49.0
YOLOv5l	640	49.0	67.3	430	10.1	2.7	46.5	109.1
YOLOv5x	640	50.7	68.9	766	12.1	4.8	86.7	205.7
YOLOv5n6	1280	36.0	54.4	153	8.1	2.1	3.2	4.6
YOLOv5s6	1280	44.8	63.7	385	8.2	3.6	12.6	16.8
YOLOv5m6	1280	51.3	69.3	887	11.1	6.8	35.7	50.0
YOLOv5l6	1280	53.7	71.3	1784	15.8	10.5	76.8	111.4
YOLOv5x6 + TTA	1536	55.0	72.7	3136	26.2	19.4	140.7	209.8



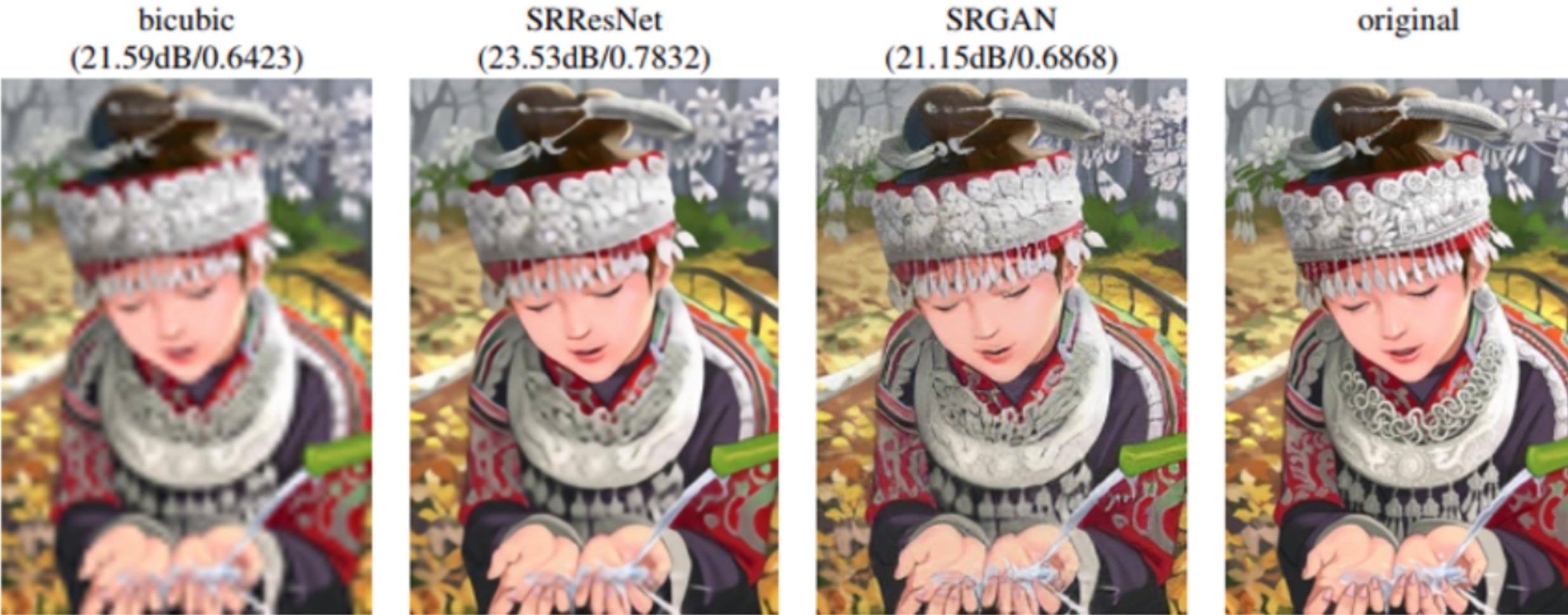
02. About Model

- Super Resolution

" **SwinIR** "

02. About Model

- Super Resolution



02. About Model

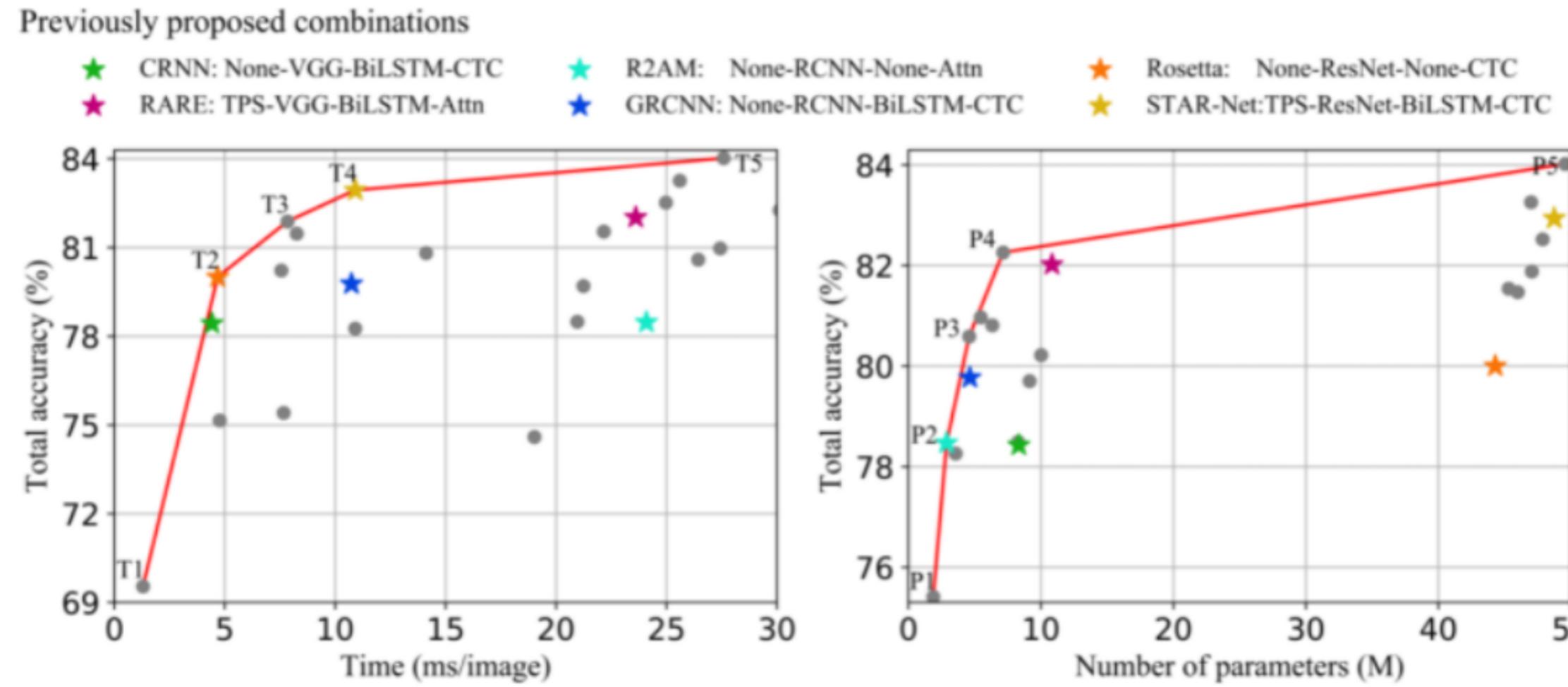
- Recognizer

" **STAR-Net** "

" **SVTR** "

02. About Model

- Recognizer : STAR-Net



02. About Model

- Recognizer : STAR-Net

탐지된 이미지의 수 : 37

Text Detection Spend Time : 0.68253 sec

Text Recognition Spend Time : 1.37478 sec

02. About Model

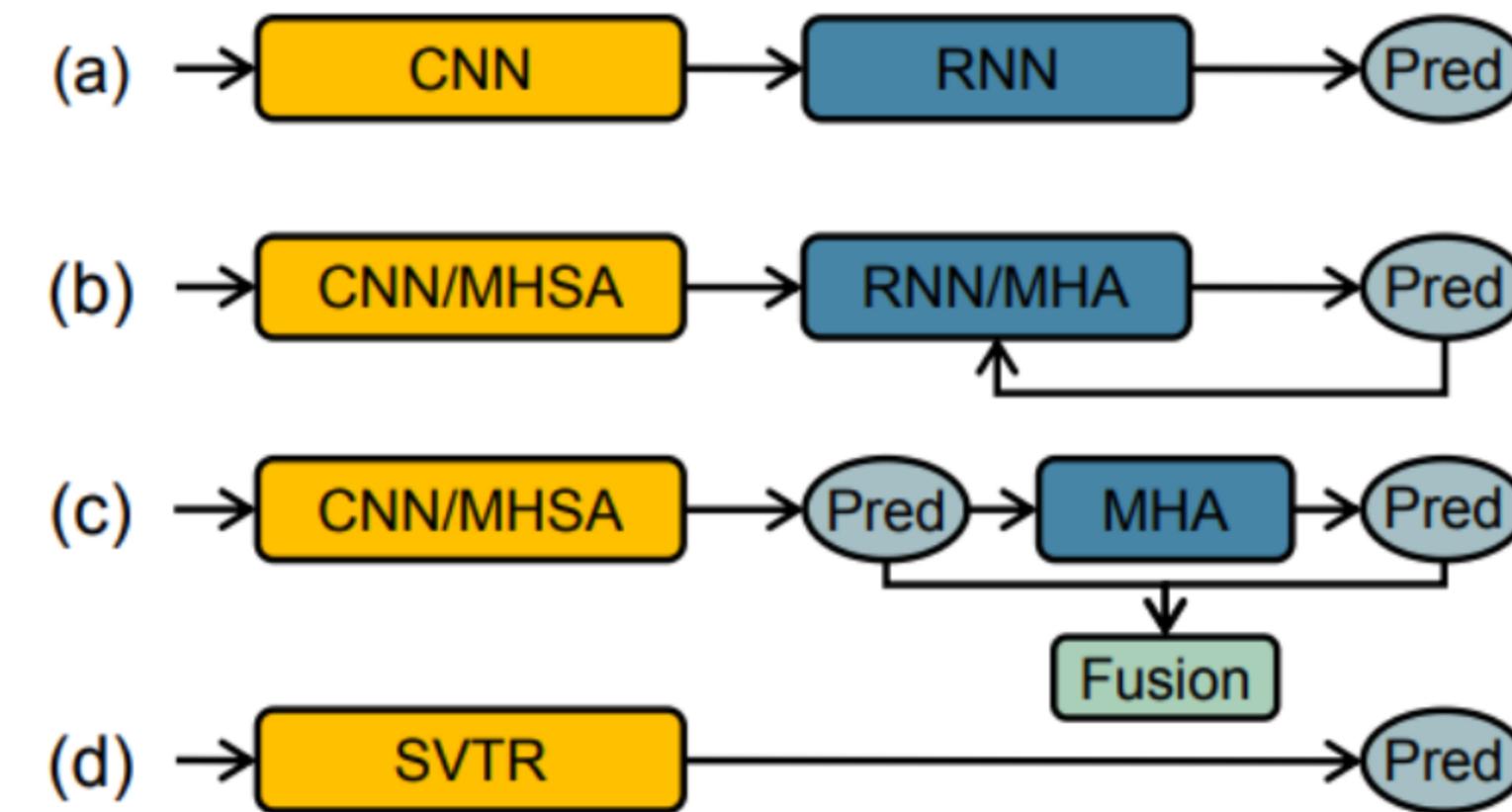
- Recognizer

" STAR-Net "

" SVTR "

02. About Model

- Recognizer : SVTR



02. About Model

- Recognizer : SVTR

탐지된 이미지의 수 : 37

Text Detection Spend Time : 0.74732 sec

Text Recognition Spend Time : 0.09569 sec

* STAR-Net 약 14배 속도 차이

02. About Model

- Recognizer

Model	Dataset	epochs	acc	norm_ED	FPS
STARNet ¹	Textinthewild(22만개)	30	93.48	96.91	180
STARNet	Augmented Textinthewild (36만개)	20	93.03	96.61	180
STARNet ²	Augmented Textinthewild (36만개)	30+25	94.00	97.07	180
SVTR_tiny ³	Textinthewild(22만개)	40	89.28	94.83	540
SVTR_tiny	Augmented Textinthewild (36만개)	40	89.86	95.19	540
SVTR_large	Textinthewild(22만개)	20	91.03	95.83	140

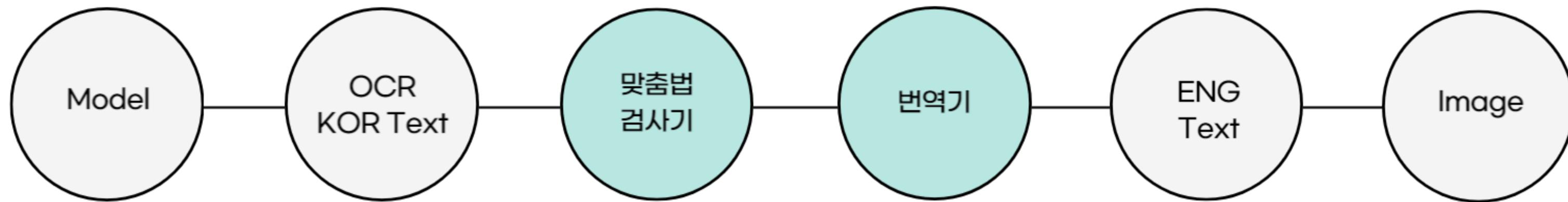
¹ STARNet은 iteration 기준으로 학습해서 epochs을 일정하게 맞추지 못했음

² 1번의 weigh를 pretrained model로 사용해서 학습

³ STARNet의 fps는 180, SVTR_tiny는 540, SVTR_large는 140으로 이에 따라 inference, train 속도에 차이가 있었음

02. About Model

- Post-Processing



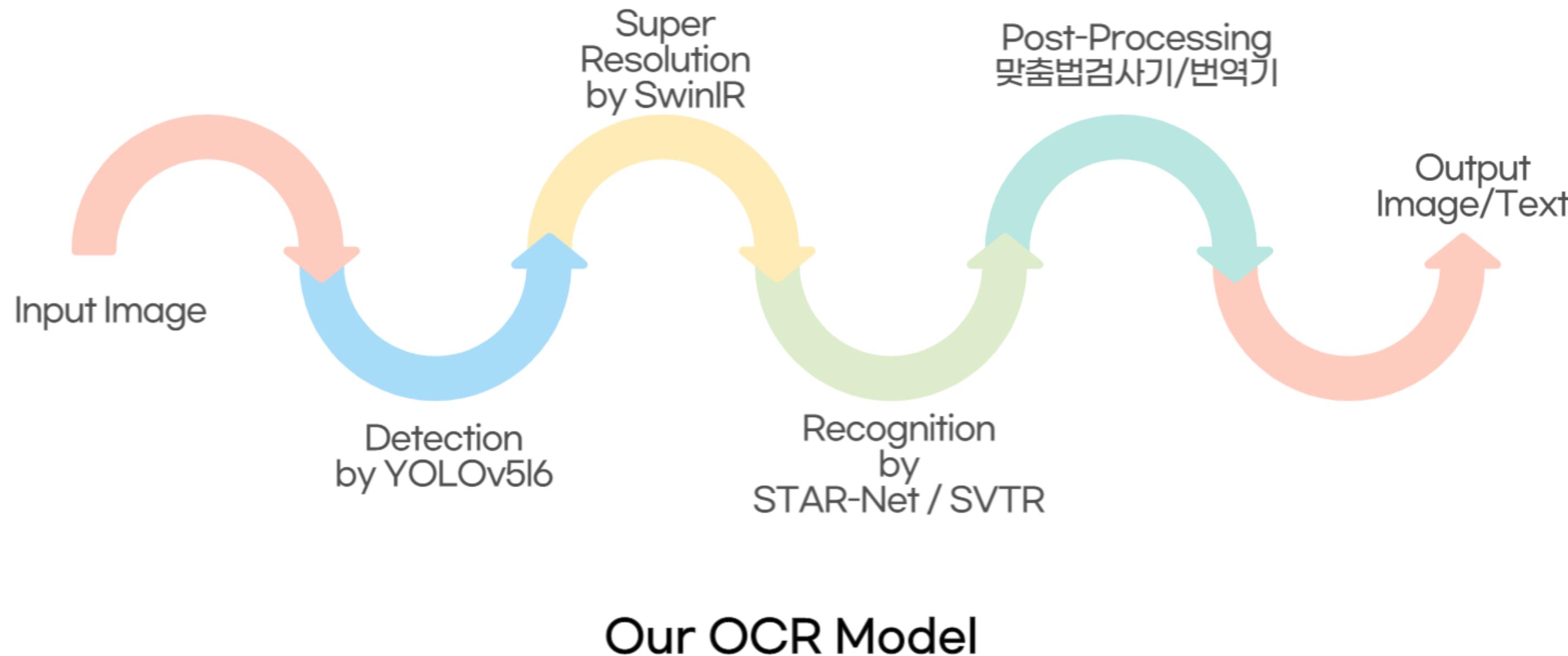
03

Preview Demo Model

- Structure
- Demonstration

03. Preview Demo Model

- Structure



03. Preview Demo Model

- Demonstration : Detection

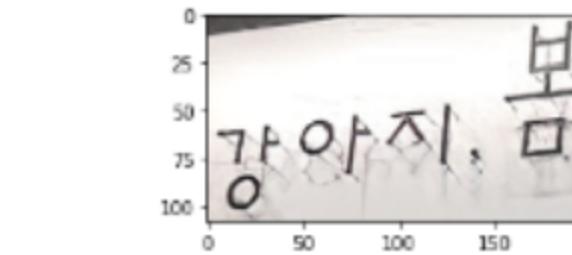
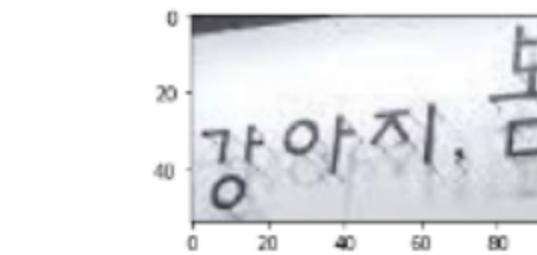
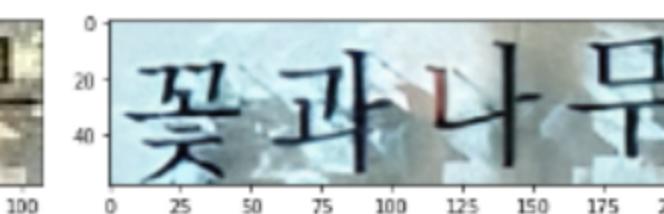
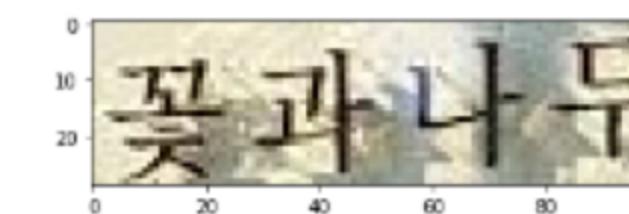
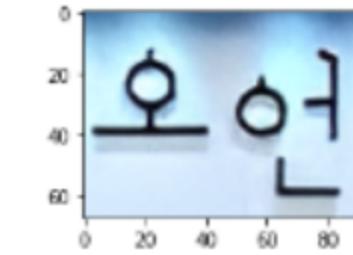
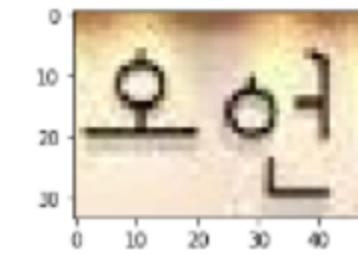
한식마당		강원도 감자 옹심이	
김치찌개	7,000	오늘백반	5,000
된장찌개	7,000	오징어백반	6,500
부대찌개	7,000	제육백반	7,500
동태찌개	7,000	낙지백반	8,000
순두부찌개	7,000	우렁쌈밥	8,000
고추장찌개	7,000	소주 4,000 청하 4,000	
청국장	7,500	맥주 4,000 복분자 10,000	
뚝배기불고기	8,000	막걸리 3,000 음료수 2,000	

메인메뉴		계절메뉴	
옹심이	8,000	서리태콩국수	8,000
옹심이 칼국수	7,000	콩나물밥	8,000
함박스테이크	9,000	톳밥	8,000
낙지덮밥	8,000	비빔국수	7,000
비빔막국수	7,000	냉모밀	7,000
메밀전병	8,000	주류·음료	
감자전	8,000	맥주, 소주 4,000	
		막걸리 3,000	
		음료 2,000	

< [원산지] 소고기:뉴질랜드산 / 돼지고기: 국내산 / 쌀,보리,배추김치:국내산 / 고추가루:국내산 / 낙지:중국산 / 서리태콩:중국산 >

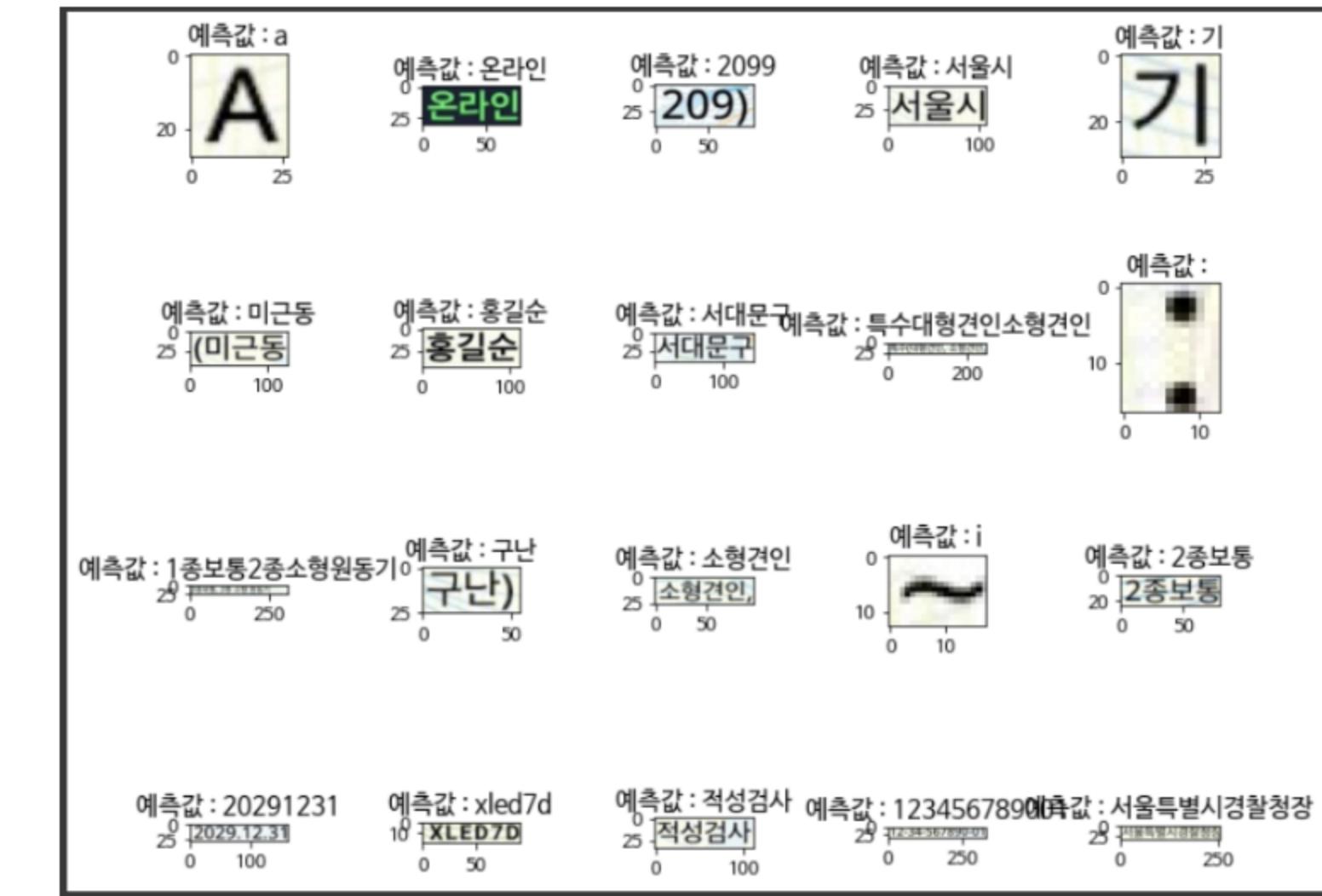
03. Preview Demo Model

- Demonstration : Super Resolution



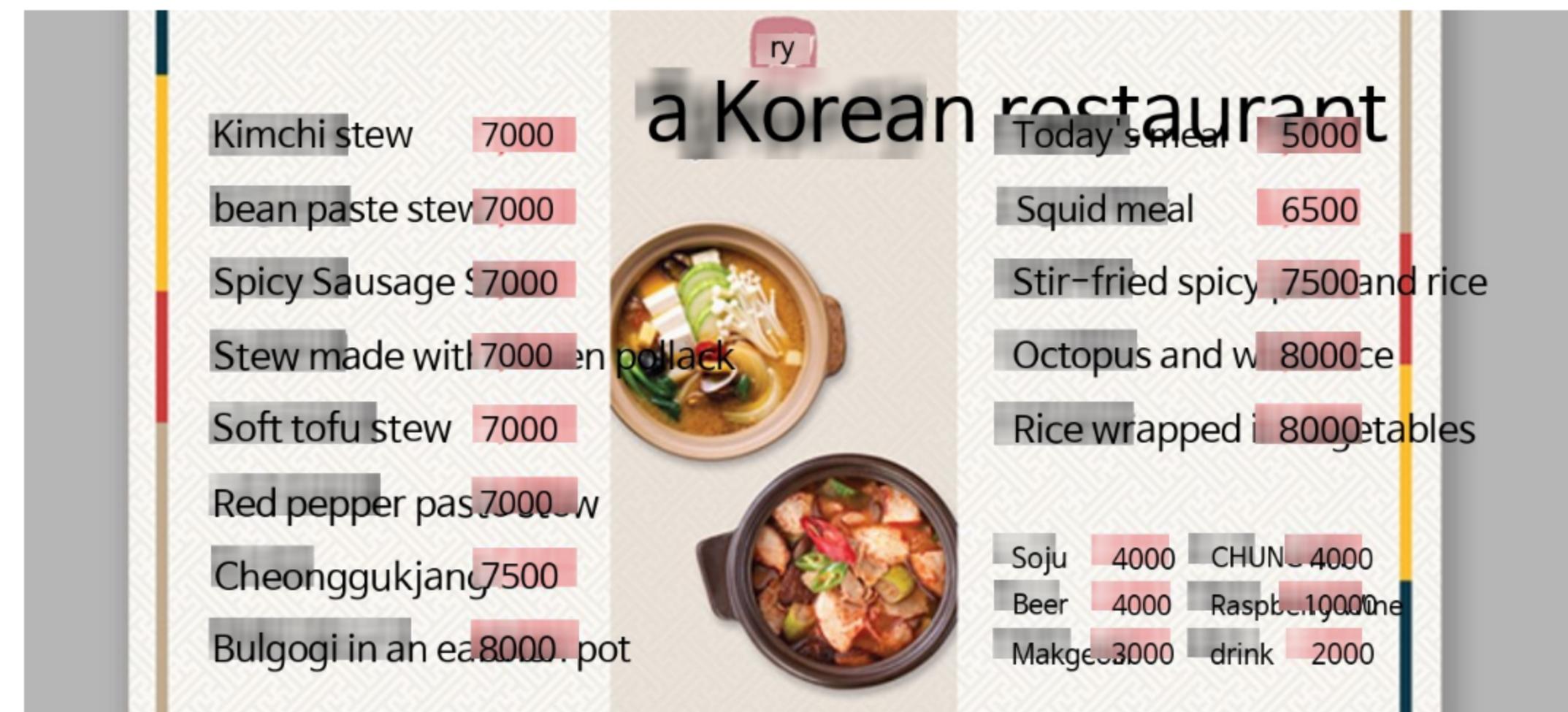
03. Preview Demo Model

- Demonstration : Recognition



03. Preview Demo Model

- Demonstration : Output



04

Difficulties

- Detection
- Super Resolution
- Recognition
- Post-Processing

04. Difficulties

- Detection

01 한글 detection 성능 관련

- 학습된 가중치를 기준으로 이미지를 detection 했을 시 성능이 떨어짐
- inference setting을 통해 conf, iou 값을 낮춤

03 Detection 범위 문제

- threshold 및 inference setting을 해도 한계가 있음
- Text snake 모델 사용 토의 : pre-trained 값 사용

02 inference setting 변경하지 않기 위해 새로운 학습을 시도

- Augmentation : rotation, shear 등
- Background Image 추가

04 모델 비교 시, 데이터셋을 통일

04. Difficulties

02 inference setting 변경하지 않기 위해 새로운 학습을 시도

Original

meters, 0 gradients, 109.9 GFLOPs	P	R	mAP@.5	mAP@.5:.95:
	0.927	0.906	0.966	0.8

Augmentation

meters, 0 gradients, 109.9 GFLOPs	P	R	mAP@.5	mAP@.5:.95:
	0.938	0.924	0.973	0.814

Background image

meters, 0 gradients, 109.9 GFLOPs	P	R	mAP@.5	mAP@.5:.95:
	0.926	0.908	0.966	0.8

04. Difficulties

03 Detection 범위 문제



BBox



Text Snake

04. Difficulties

- Super Resolution

속도 문제

해상도가 높은 사진의 경우엔 SR 시간이 오래걸리게 됨

4배 확대에서 2배 확대버전으로 변경

Crop Image(Bounding Box)의 크기가 우리가 선택한
기준보다 낮을 때만 적용하는 것으로 변화

04. Difficulties

- Recognition

영어 Test Dataset에 대한 정확도 감소

- 모델 정확도 87% -> 82%

학습한 Dataset과 비슷한 경우에만 정확도 높음

- 학습한 dataset에 따라 다른 유형의 image를 예측할 경우 정확도가 매우 낮음
: ex) 공공문서 학습한 model 경우, 간판 이미지 예측 정확도가 낮음

두줄로 된 Image, 학습하지 않은 특수문자 등은 인식하지 못함

04. Difficulties

- Post-Processing

번역 문제

- ▶ 한 글자씩 인식되는 경우 번역의 문제

맞춤법 문제

- ▶ 고유 명사, 단어 생략된 경우 맞춤법 검사 오류 발생



공부 잡하는 부원이



공부 잘하는 부원이

05

Conclusion

- Future map
- Q&A

05. Conclusion

- Future Map

메뉴판, 영수증, 신용카드, 고지서,
명함, 신분증, 공문서 등

TTS

번역기

문서 요약

OMR 정답지 인식

모델 최적화

05. Conclusion

- Q&A

01

Introduction

- About OCR
- Advanced Work
- Project objective

02

About Model

- Detector
- Super Resolution
- Recognizer
- Post-Processing

03

Our OCR Model

- Structure
- Demonstration

04

Limitations & Solve

- Detection
- Super Resolution
- Recognition
- Post-Processing

05

Conclusion

- Future Directions
- Q&A

코드득 (Code Gain)

발표를 들어주셔서

감사합니다