

# Highlight Text Extraction Pipeline

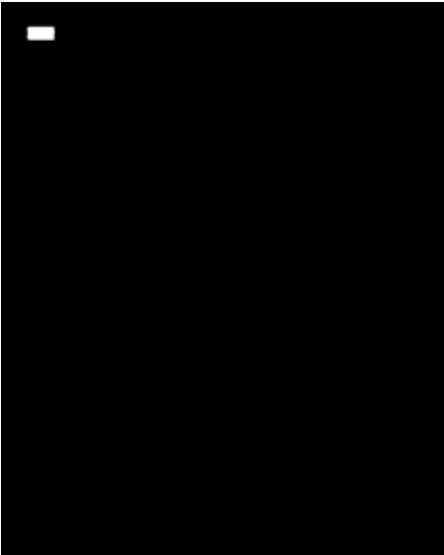
## Original → Detection → OCR

(a) Original Image  
(Yellow Highlight)

입성 데이터 생성은 리벨링 비용을 절감하고 대규모 데이터셋을 구축합니다.



(b) Color Detection  
(HSV Mask)



(c) OCR Extraction  
(Tesseract)

Detected: "입성"

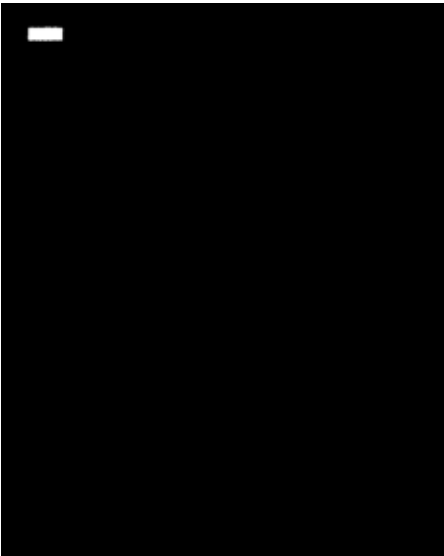
데이터 생성은 리벨링 비용을 절감하고 대규모 데이터셋을 구축합니다.

(d) Original Image  
(Green Highlight)

이미지 전처리는 노이즈 제거, 이진화, 대미 향상 등을 포함합니다.



(e) Color Detection  
(HSV Mask)



(f) OCR Extraction  
(Tesseract)

Detected: "입성"

데이터 생성은 리벨링 비용을 절감하고 대규모 데이터셋을 구축합니다.