

8주차

팀원: 송성근, 황동욱, 조선빈

```
lookup.KeyValue  
f.constant(['em  
=tf.constant([0  
ce = tf.lookup.StaticV  
init,  
num_oov_buckets=5)
```

```
lookup.StaticVocabular  
initializer,  
num_oov_buckets,  
lookup_key_dtype=None  
name=None,  
experimental_is_sparse
```

구조

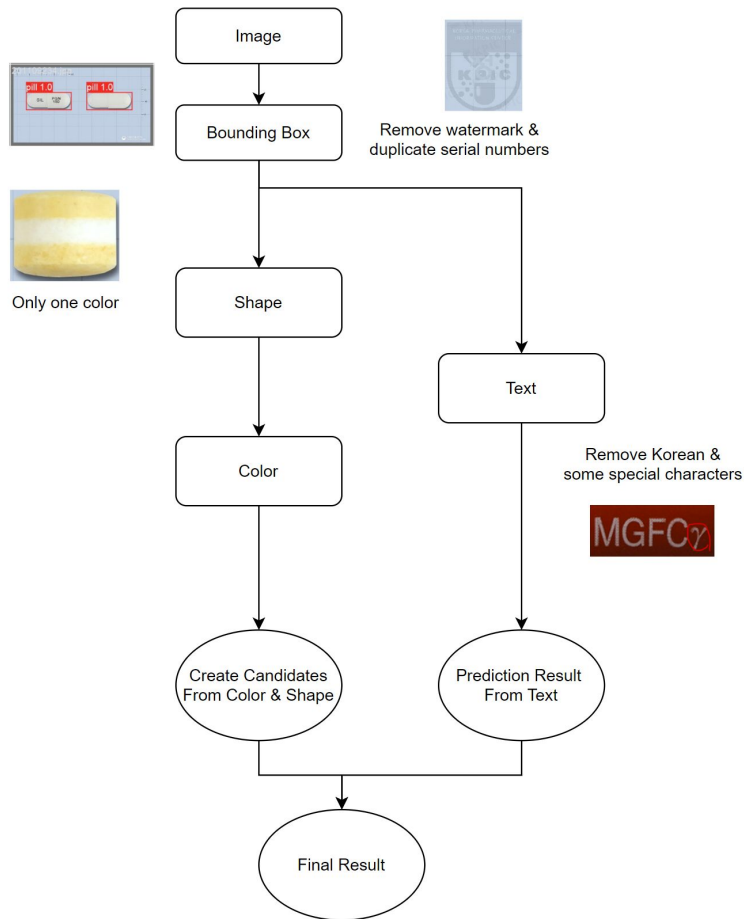


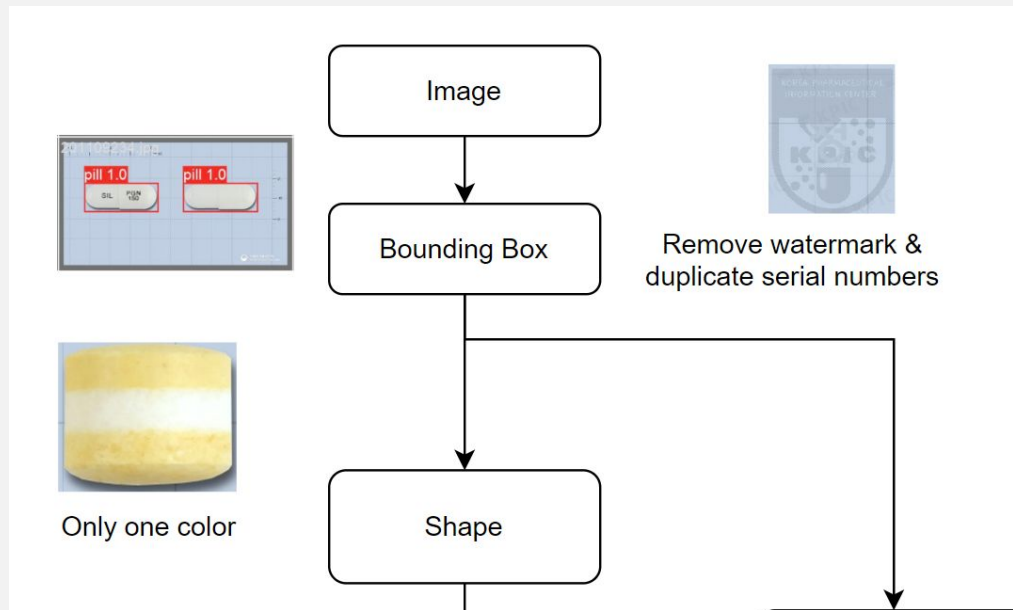
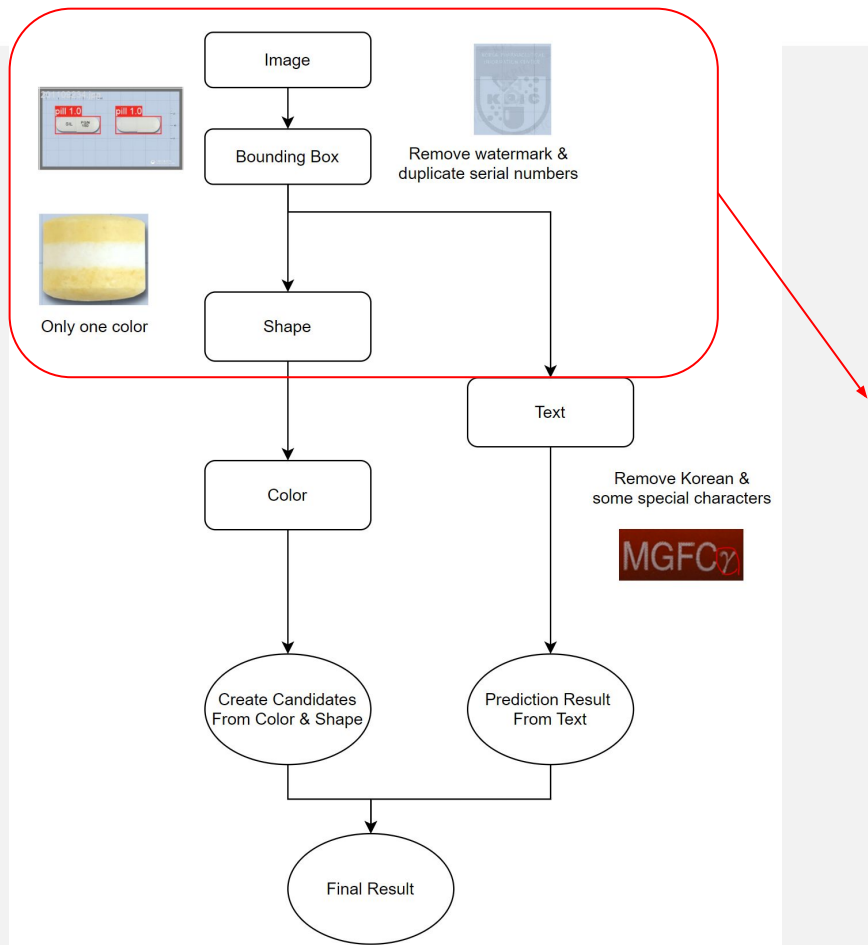
Image > Bounding Box

: 워터마크/시리얼넘버 중복 등의 문제가 있는 것들은 제거

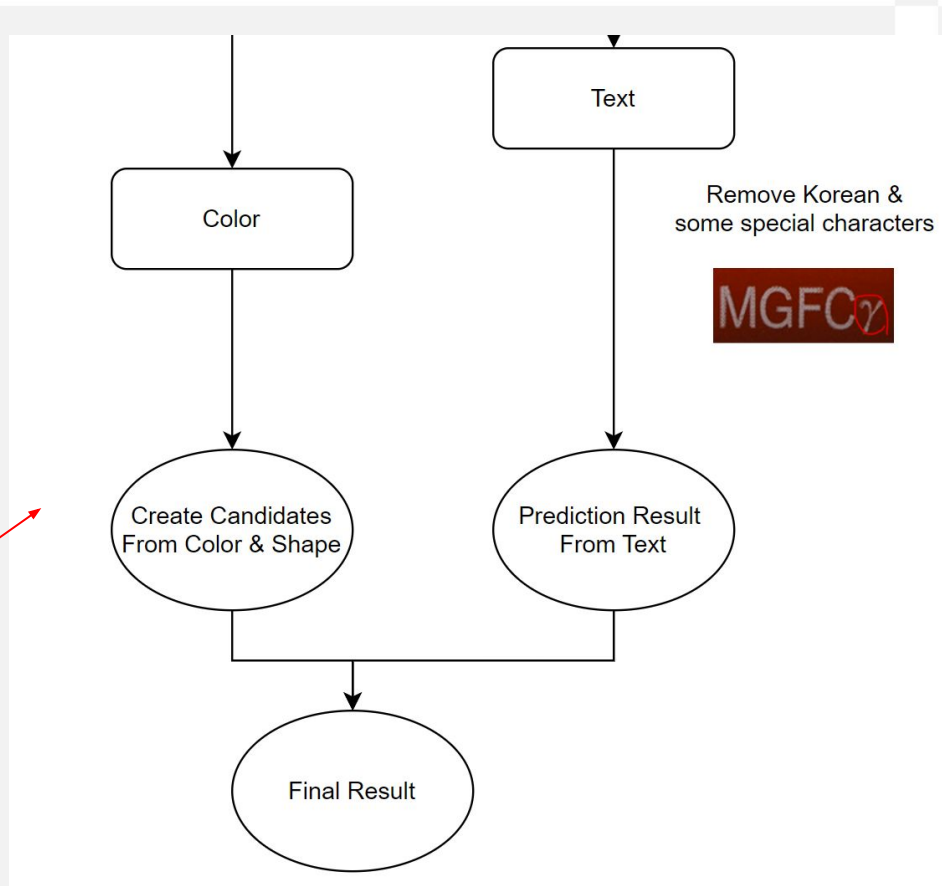
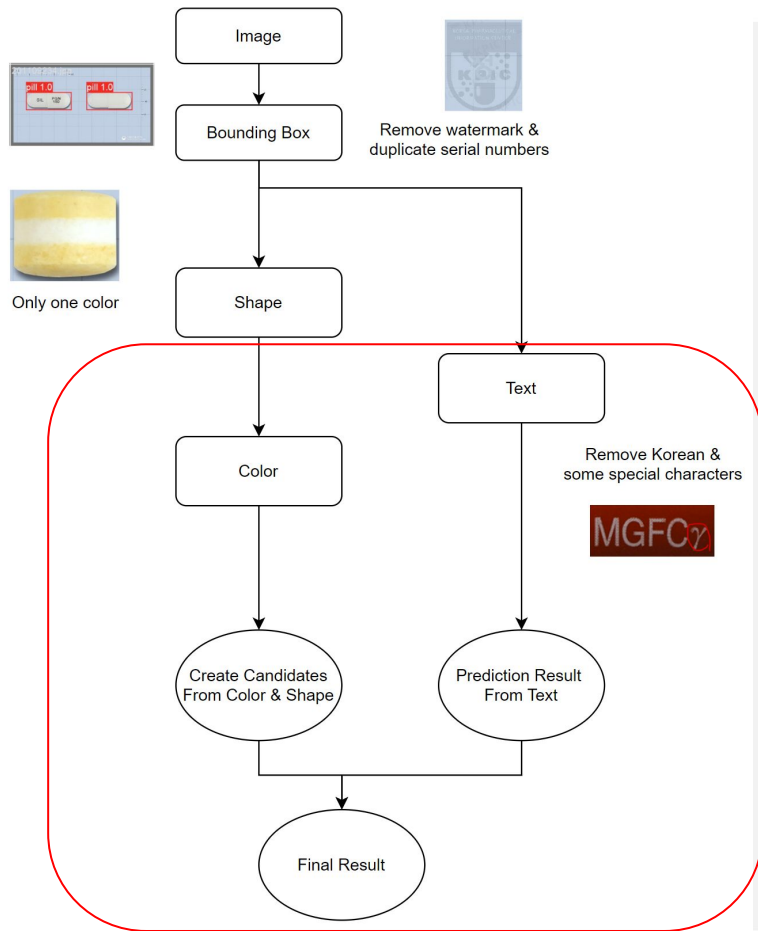
알약 중 두개 이상의 색이 있는 알약들도 있는데, 처리하기 어려움이 있어서 일단 제거하고 한가지 색만 있는 알약으로 진행 후에 시간이 된다면 추가해볼 예정

텍스트에서 특수문자나 한글은 잘 인식되지 않아서 해당 문자들을 포함한 알약들은 일단 제거하고 진행 후에 시간이 된다면 추가해볼 예정

구조



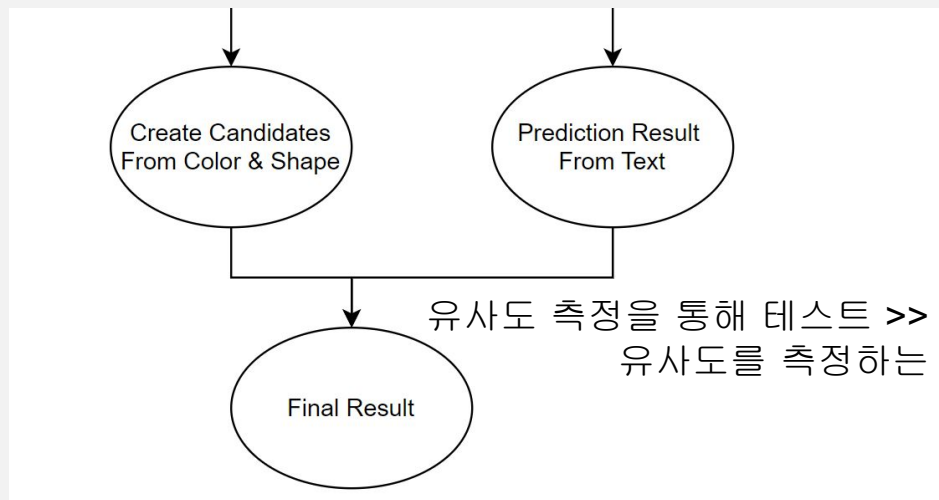
구조



유사도 측정

Jaccard Similarity : 두 집합 간의 교집합과 합집합의 비율을 계산하여 유사도를 측정하는 방식

Levenshtein Distance : 두 텍스트를 서로 같게 만들기 위해 필요한 최소 편집(삽입, 삭제, 대체)



유사도 측정을 통해 테스트 >> 결과가 생각보다 좋지 않았음
유사도를 측정하는 좋은 알고리즘을 탐색해서 적용해볼 예정