

딥러닝 및 다중 시각 특징 분석을 이용한 자동 경구약제 식별 시스템

TEAM 8

김민정 (2391007)

이수안 (2391016)

이윤지 (2391017)

Table of Contents

- 연구 배경 및 필요성
- 프로젝트 목적 및 가설
- 구현 과정 및 방법
 - YOLO 기반 다중 알약 탐지
 - ResNet 기반 알약 특징(색상, 형태) 추출
 - 분할선 검출
 - 약제 각인 텍스트 추출 과정(OCR)
- 기대 효과 및 개선점

연구 배경 및 필요성

- 다수의 약물을 장기간 복용하는 **다중 약물 복용 계층** 급증
 - 현재 임상 현장과 가정에서의 약물 식별은 약사, 환자에 의존하는 수동적인 방식
- 약사의 조제 검수 과정에서 **과도한 인지적 부하**를 유발하고,
업무 효율을 저하시키며 피로 누적으로 인한 인적 오류의 원인이 되는 경우도 존재
- 시각적 인지 능력이 저하된 노인 및 시각장애인은 **심각한 오복용 위험**에 노출되어 있음

복잡한 시각적 환경에서 약물을 신속하고
정확하게 식별할 수 있는 **자동화된 보조 시스템 개발의 필요성**

선행 연구 검토 및 문제 제기

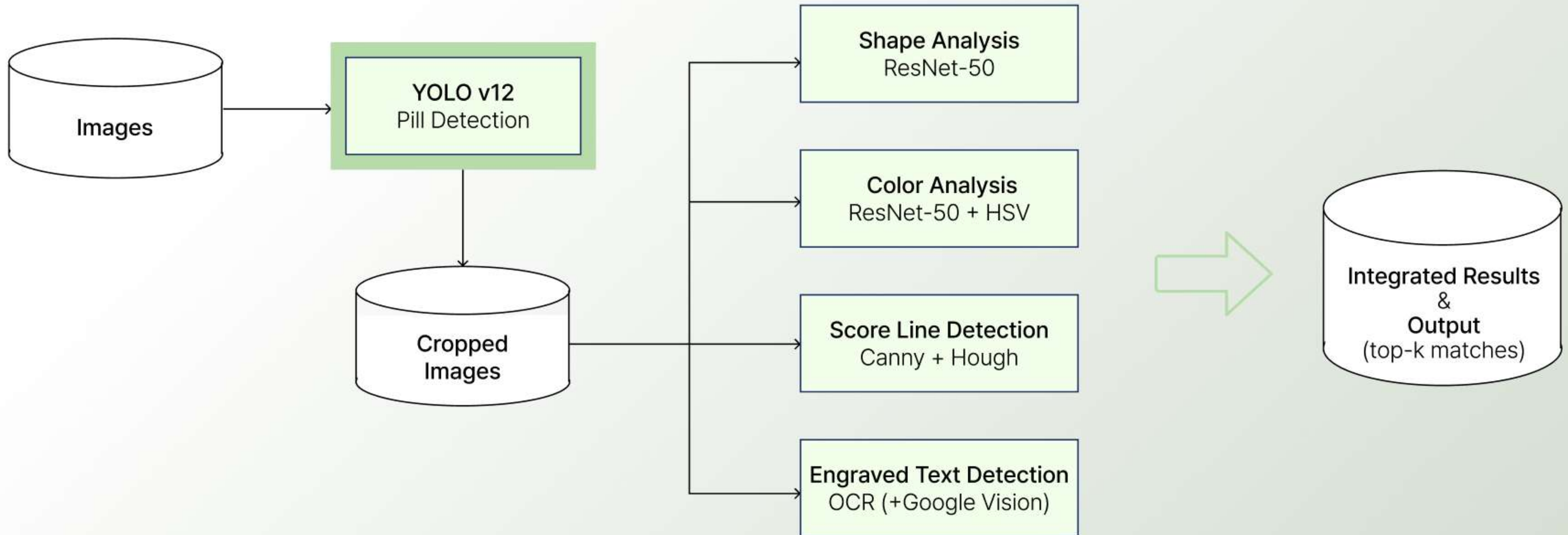
- 초기 연구들은 전통적인 이미지 처리 기법을 활용하여 알약의 윤곽선과 색상 히스토그램을 분석하며 약물 식별을 수행하였음 (Caban et al., 2012)
- 이후 합성곱 신경망(CNN)을 비롯한 딥러닝 기반 접근이 도입되어 복잡한 시각적 변형에도 robust한 약물 인식 모델이 제안되었음 (Zeng et al., 2017)
- 대부분의 기존 연구들은 단일 객체 분류에 기반한 접근을 취하며, 흰색 배경 위에 알약 한 알이 놓인 이상적인 실험 환경을 가정하는 경우가 많음
- 여러 종류의 알약이 무질서하게 섞이거나 겹쳐지고, 다양한 조명과 그림자가 존재하는 복용 환경에서는 단순 분류 기반 모델의 성능이 현저히 저하되는 한계가 있음

프로젝트 목적 및 가설

기존 단일 객체 분류 연구의 한계를 넘어, 혼합된 알약 이미지로부터 개별 약제를 정밀하게 탐지하고 식별하는 통합 딥러닝 파이프라인을 제안하고 검증하는 것

- **첫째**, 객체 탐지 알고리즘인 YOLO(You Only Look Once)를 적용할 경우, 알약이 겹치거나 흐뿌려진 다중 객체 환경에서도 90% 이상의 높은 재현율(Recall)로 개별 알약의 관심 영역(ROI)을 분리할 수 있을 것이다.
- **둘째**, 알약 식별 단계에서 단일 특징(색상 또는 모양)만을 사용하는 것보다 형상, 색상, 분할선 및 각인(OCR) 정보를 결합한 **다중 특징 융합 방식**을 적용할 때 식별 정확도가 유의미하게 향상될 것이다.
- **셋째**, 알약 표면의 각인은 양·음각 형태로 시인성이 낮아 단독 인식률은 낮을 수 있으나, 이를 시각적 특징(모양, 색상)과 **결합하여 필터링 조건으로 활용**한다면 유사 약제 간의 오분류를 효과적으로 감소시킬 수 있을 것이다.

전체 파이프라인



데이터 수집 및 구성 - 다중 객체 탐지

다중 객체 탐지와 개별 약제 식별이라는 상이한 두 과제를 수행하기 위해,
각 목적에 최적화된 별도의 데이터셋을 구축하여 활용하였음

Roboflow Universe의 공개 데이터셋

Roboflow 100 pills

+

seblful Pills Detection

조명 변화, 그림자, 알약의 겹침 등 실생활에서
발생할 수 있는 다양한 노이즈 환경을 포함함

AI Hub의 '경구약제 이미지 데이터'

국내 다빈도 경구약제 5,000종
(전문의약품 3,000종, 일반의약품 2,000종)

YOLO 형식으로 변환

- 데이터 누수(Data Leakage)를 방지하기 위해
이미지를 무작위로 섞는 대신,
- 원본 촬영 세트(Source Group) 단위로 Train/
Validation/Test 셋을 분할

YOLO 기반 다중 알약 탐지

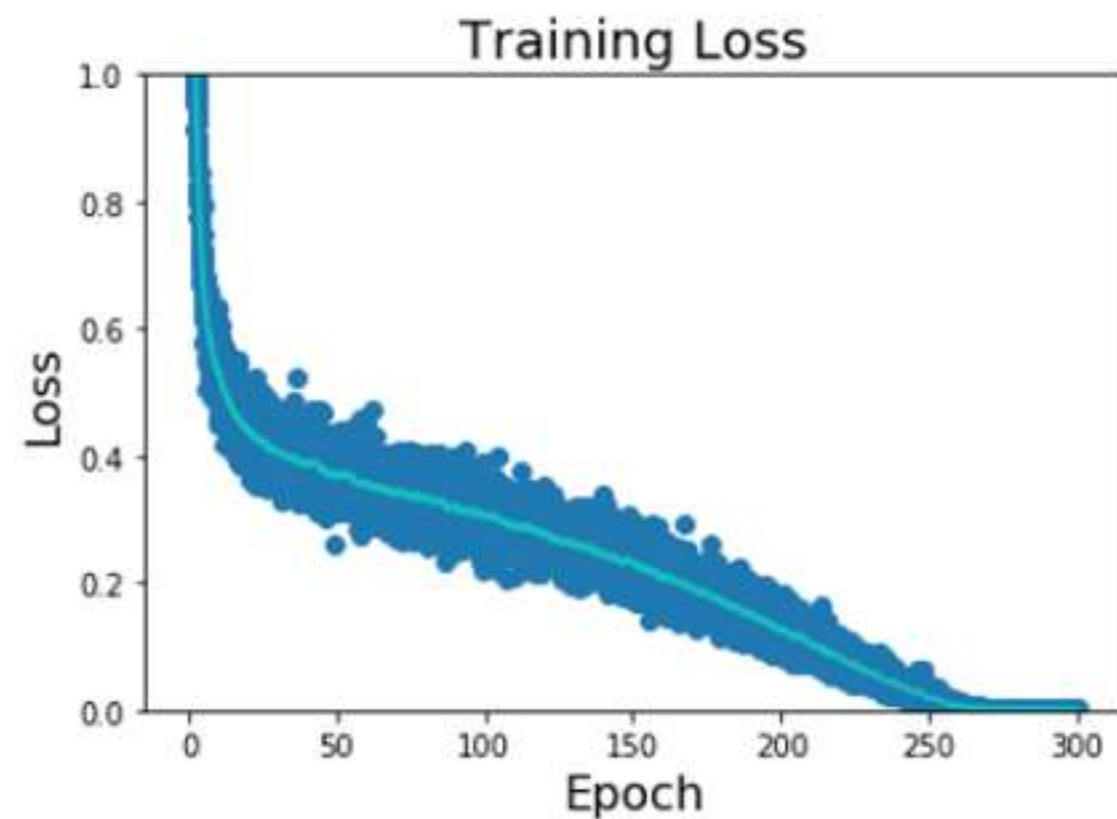
YOLOv12: Attention-Centric Real-Time Object Detectors

Yunjie Tian
University at Buffalo
yunjiet@buffalo.edu

Qixiang Ye
University of Chinese Academy of Sciences
qxye@ucas.ac.cn

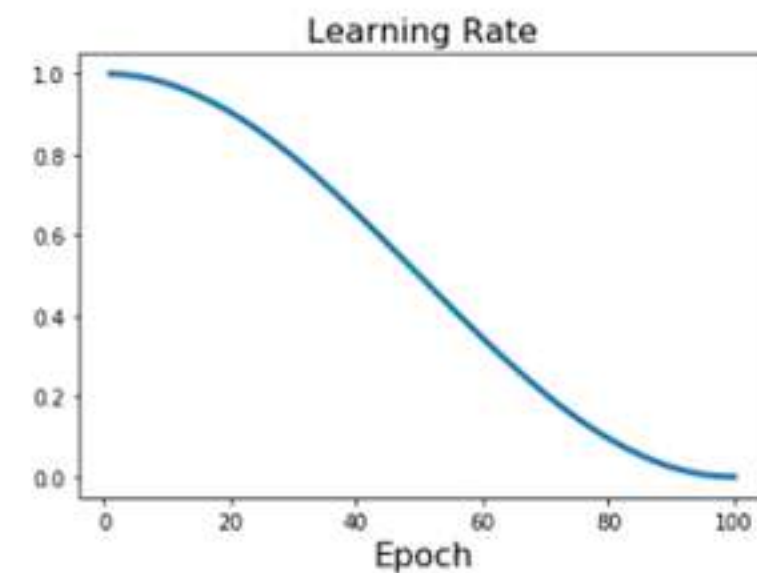
David Doermann
University at Buffalo
doermann@buffalo.edu

Learning Rate Decay: Cosine



Step: Reduce learning rate at a few fixed points.
E.g. for ResNets, multiply LR by 0.1 after epochs 30, 60, and 90.

Cosine: $\alpha_t = \frac{1}{2}\alpha_0 (1 + \cos(t\pi/T))$



Loshchilov and Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts", ICLR 2017
Radford et al, "Improving Language Understanding by Generative Pre-Training", 2018
Feichtenhofer et al, "SlowFast Networks for Video Recognition", ICCV 2019
Radosavovic et al, "On Network Design Spaces for Visual Recognition", ICCV 2019
Child et al, "Generating Long Sequences with Sparse Transformers", arXiv 2019

- Ultralytics YOLO 모델 중 yolo12n을 채택
- 소형 객체인 알약의 특징 손실을 방지하기 위해 입력 이미지 크기(imgsz)를 960px로 설정하여 리사이즈
- 학습은 총 200 Epochs, 배치 크기는 8
- 최적화 함수로는 AdamW를 사용
- Cosine Decay 스케줄러를 적용하여 학습률을 조절
- 과적합 방지를 위해 30 Epochs 동안 성능 개선이 없을 경우 학습을 조기 종료

YOLO 기반 다중 알약 탐지

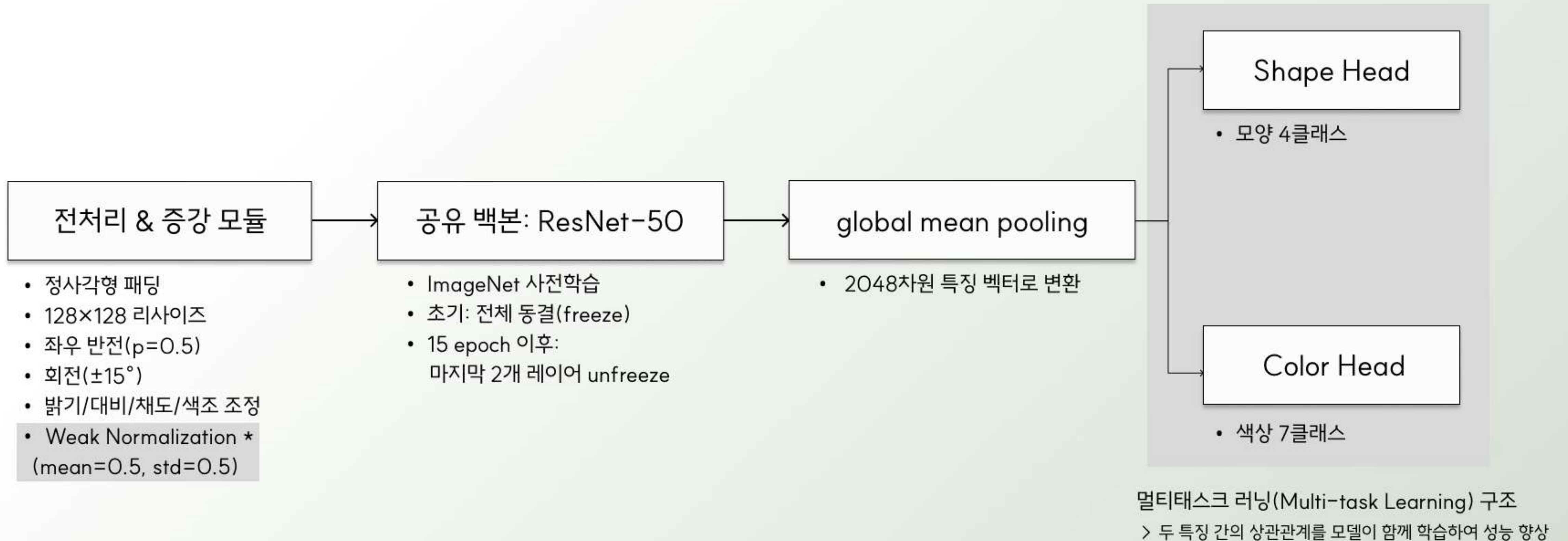
데이터의 절대적 수가 제한되어 있고, 실제 환경에서 발생할 수 있는 다양한 상황에 대응하기 위해 복합적인 데이터 증강 기법을 적용

```
hsv_h: float = 0.02 # 0.015 → 0.02 (색조 변화 증가)
hsv_s: float = 0.8 # 0.7 → 0.8 (채도 변화 증가)
hsv_v: float = 0.5 # 0.4 → 0.5 (명도 변화 증가)
degrees: float = 15.0 # 0.0 → 15.0 (회전 추가)
translate: float = 0.15 # 0.1 → 0.15 (이동 증가)
scale: float = 0.6 # 0.5 → 0.6 (크기 변화 증가)
shear: float = 2.0 # 0.0 → 2.0 (전단 변환 추가)
perspective: float = 0.0001 # 0.0 → 0.0001 (약간의 원근 변환)
flipud: float = 0.5 # 상하 반전
fliplr: float = 0.5 # 좌우 반전
mosaic: float = 1.0 # Mosaic augmentation 활성화
mixup: float = 0.15 # 0.0 → 0.15 (Mixup 추가)
copy_paste: float = 0.1 # Copy-paste augmentation 추가
```

Data Augmentation 설정

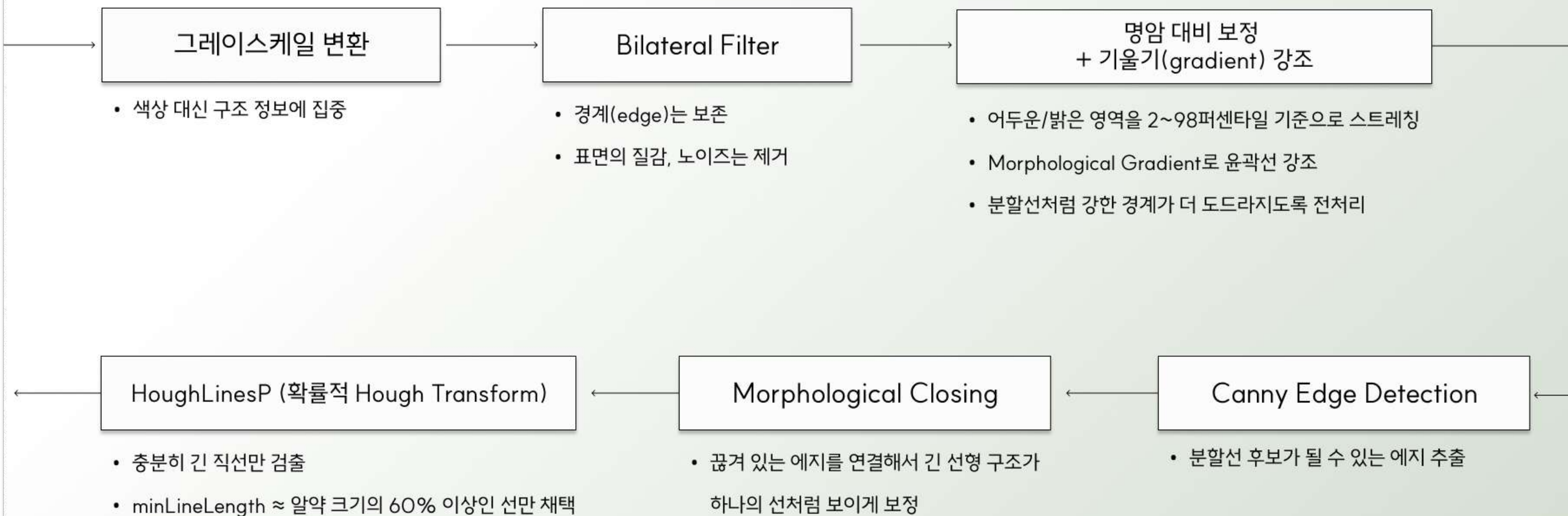
- 색상·밝기 변화 (HSV 기반 Color Jitter)
- 회전, 이동, 스케일, Shear, 원근 변형
- 상하/좌우 Flip
- Mosaic, Mixup, Copy-Paste 고급 증강
- 마지막 10 epoch는 Mosaic 비활성화
 - 실제 이미지 분포와 유사한 환경에서 fine-tune을 위함

ResNet 기반 알약 특징(색상, 형태) 추출



* ImageNet 정규화는 색상 정보를 많이 왜곡시키나,
알약은 색상이 매우 중요한 식별 단서이므로 원본 색상 정보를 최대한 보존

분할선 검출



분할선 검출

기하학적 필터링: 알약 중심 통과 여부 검사

- 알약 중심 (cx, cy) 계산
- 선과 중심점 사이 거리 d 계산
 - $d \leq \text{tolerance}$ 인 선만 분할선 정보로 유지
- 흠집·로고·각인 텍스트에서 나온 선은 대부분 탈락

방향 분류 및 최종 타입 결정

- 각도(angle)로 수평/수직 선 분리
 - 수평만 있으면 1개 분할선('-')
 - 수평+수직 있으면 십자('+')
 - 없으면 'none'

```
def _line_passes_center(self, x1: int, y1: int, x2: int, y2: int,
                        cx: int, cy: int, tolerance: int = 15) -> bool:
    numerator = abs((y2-y1)*cx - (x2-x1)*cy + x2*y1 - y2*x1)
    denominator = np.sqrt((y2-y1)**2 + (x2-x1)**2)

    if denominator == 0:
        return False

    distance = numerator / denominator
    return distance <= tolerance
```

약 #1

검출 신뢰도: 83.80%
모양: Elliptical (타원형) (신뢰도: 76.12%)
색상 (HSV): 브라운, 핑크
색상 (CNN): 브라운 (신뢰도: 38.12%)
최종 색상: 브라운, 핑크
분할선: 없음
각인: BF (신뢰도: 100.00%)

약 각인 OCR

데이터의 절대적 수가 제한되어 있고, 실제 환경에서 발생할 수 있는 다양한 상황에 대응하기 위해 복합적인 데이터 증강 기법을 적용

전처리 (CLAHE 기반 대비 향상)

곡면·광택·그림자 때문에 흐릿해진 각인 텍스트를 선명하게 만들기 위한 과정

- 전처리 파이프라인
 - 컬러 > 그레이스케일 변환
 - CLAHE 적용
 - clipLimit = 3.0, tileGridSize = 8×8
 - 이미지를 작은 타일 단위로 나누어 히스토그램 균등화
 - 밝기 변화가 심한 영역에서도 텍스트 대비 향상
- 효과
 - 반사광·그림자·색상 변화로 텍스트/배경 경계가 흐릿한 약제에서도 각인 윤곽이 더 뚜렷하게 드러나 OCR 정확도 개선

8방향 회전 OCR

- 알약 크롭 이미지를 45도 단위로 회전($0^{\circ} \sim 315^{\circ}$)
- 각 각도마다 EasyOCR 수행 (총 8번)
- 각도별 결과 중 평균 신뢰도가 가장 높은 텍스트를 최종 후보로 선택

이중 OCR Fallback 전략

- 기본 OCR: EasyOCR
 - 속도가 빠르고, 비용이 들지 않음
- Fallback OCR: Google Vision API
 - 글자를 아예 못 찾거나,
 - 신뢰도가 임계값(예: 80%) 미만일 때만 호출

DB 매칭 전략

특징 가중치 기반 스코어링

특징	설명	가중치
각인	가장 강력한 식별 단서가 됨	3.0
색상	제한된 조합이지만 시각적으로 뚜렷	2.0
모양	제형·타입 구분에 활용할 수 있음	1.5
분할선	절취선/십자선 정보	1.0

각 특징별 일치도(score)를 계산 후

가중합 점수 = $\sum(\text{특징 일치도} \times \text{가중치})$ 로 총점 산출
총점이 높은 순으로 정렬하여 상위 9개 후보 약물 제시



< 비콤씨에프정 618.6mg/병

- 약 #1

예측 색상: ['브라운', '핑크']

예측 모양: Elliptical (타원형)

각인: BF (conf=1.00)

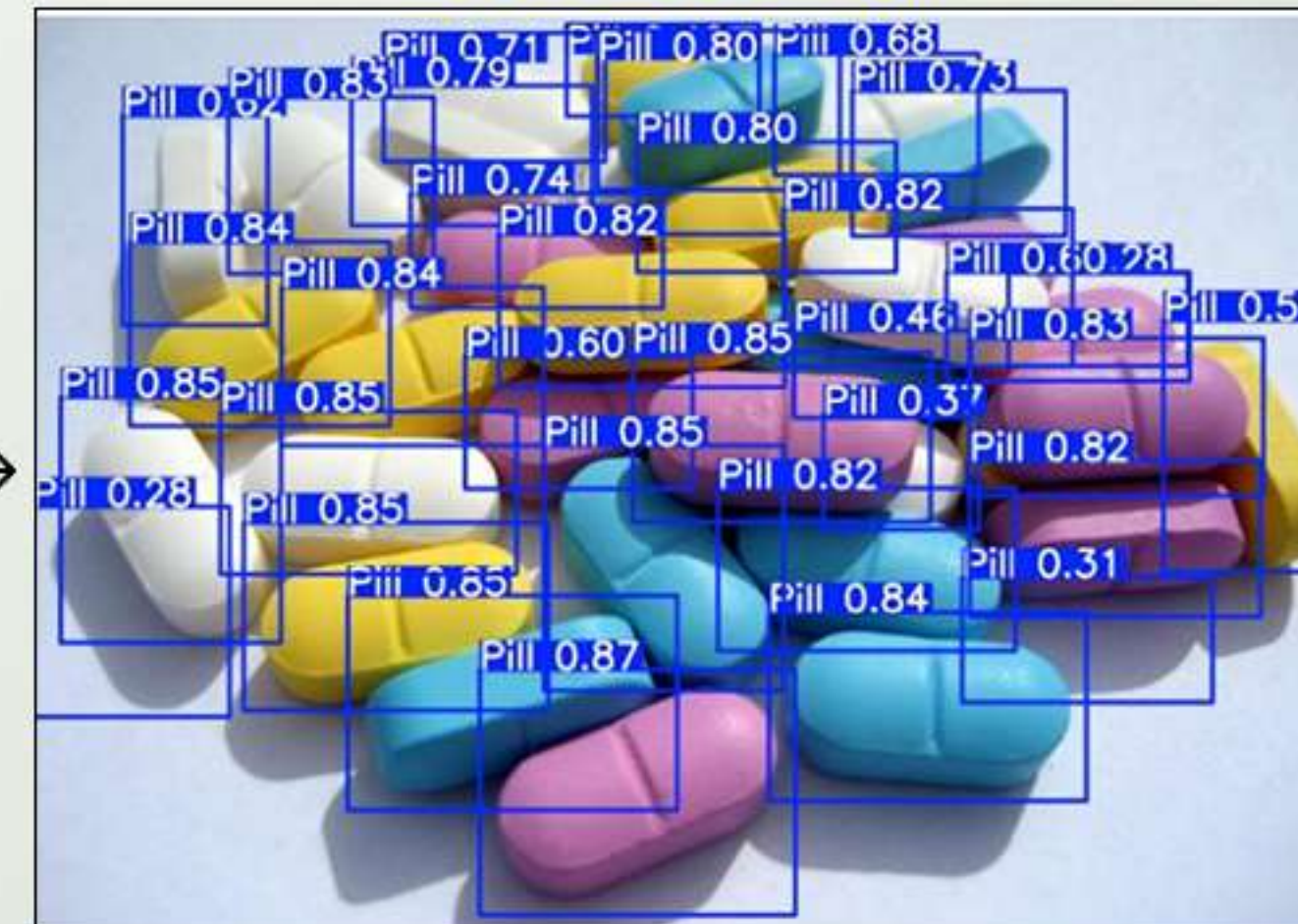
- Top 5 후보:

1. K-006192 | 비콤씨에프정 618.6mg/병 ((주)유한양행) total=5.500 [engr=1.00, color=0.50, shape=1.00, line=0.00]
2. K-004168 | 비오라민정 72.2mg/PTP (수출용) (명인제약(주)) total=5.000 [engr=0.50, color=1.00, shape=1.00, line=0.00]
3. K-030024 | 벤포메가정 50mg/병 ((유)한풍제약) total=4.500 [engr=0.67, color=0.50, shape=1.00, line=0.00]
4. K-037452 | 메가트루포커스정 215mg/병 ((주)유한양행) total=4.500 [engr=0.67, color=0.50, shape=1.00, line=0.00]
5. K-014274 | 비콤씨에이스정 10mg/PTP ((주)유한양행) total=4.357 [engr=0.29, color=1.00, shape=1.00, line=0.00]

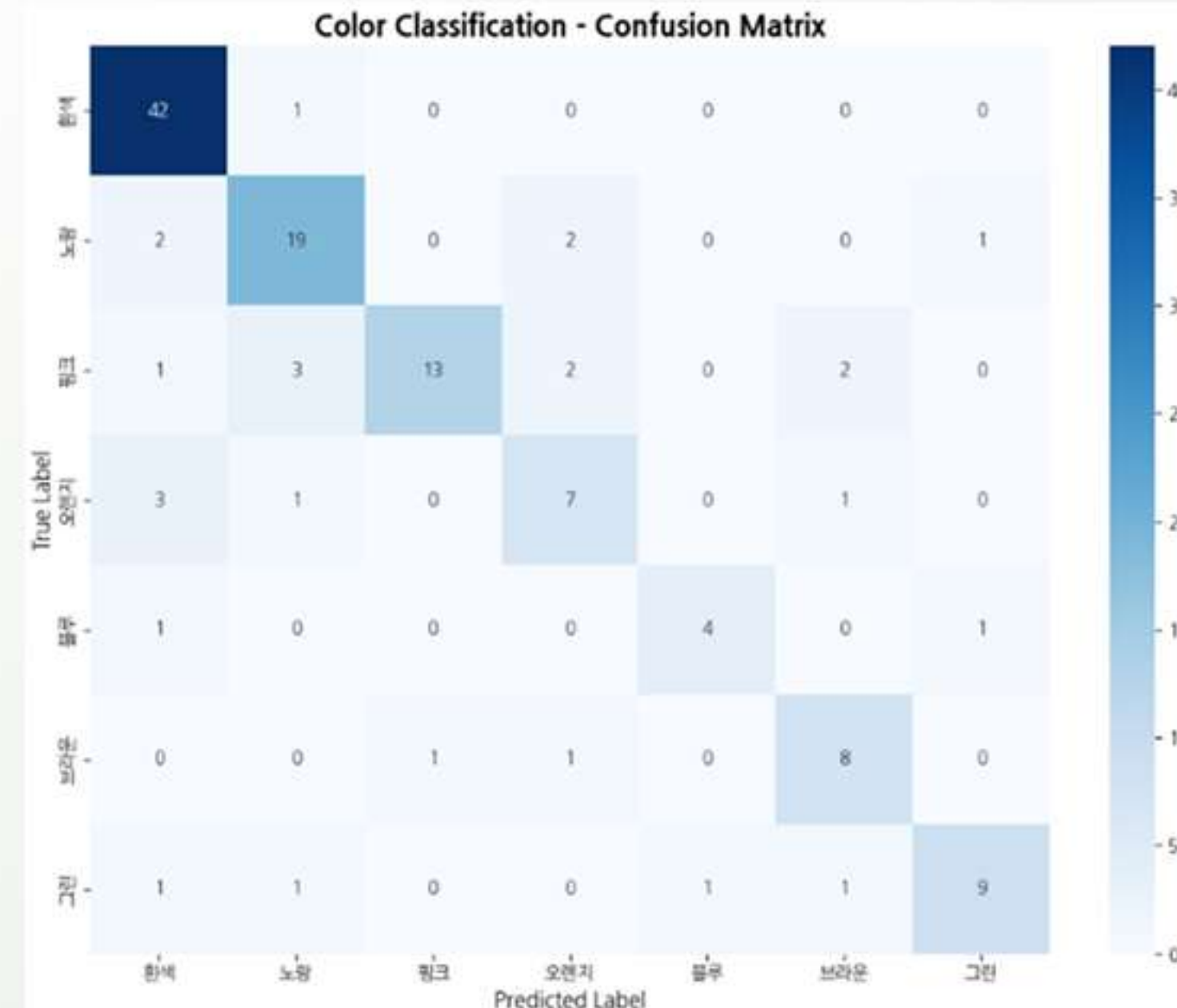
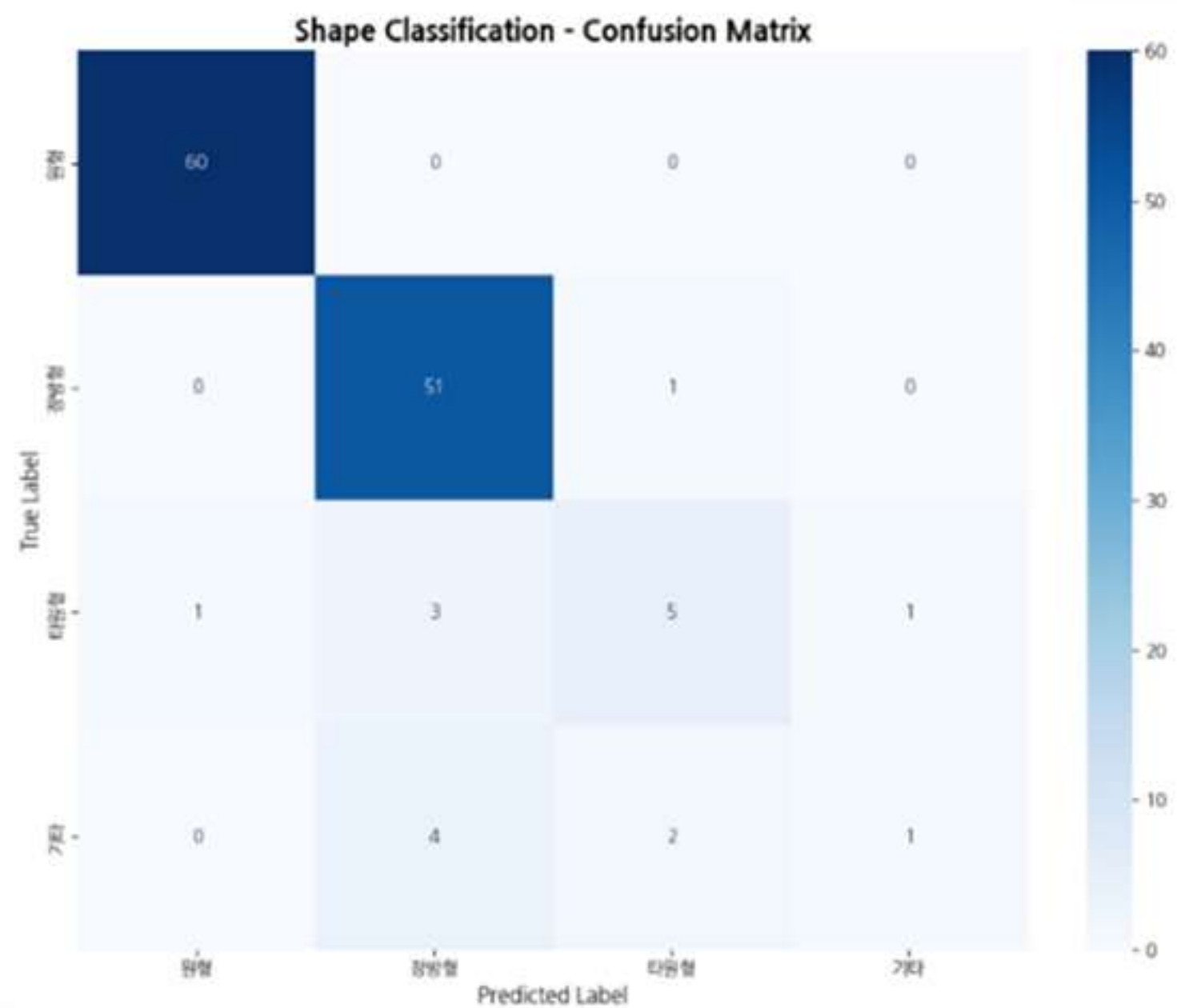
YOLO 기반 탐지 성능

Metrics	Measurement	Description
Precision	0.931	Precision of positive prediction
Recall	0.934	Object detection rate
mAP50	0.971	Mean average precision at IoU 0.5
mAP50-95	0.620	Mean average precision over IoU 0.5–0.95 thresholds

- mAP50-95에서, IoU 조건이 높아질수록 박스 위치 정확도 요구가 엄격해지기 때문에 다소 낮게 나타났으나 목표가 알약 영역(ROI) 추출인 점을 고려할 때, mAP50 수준의 성능이면 실제 활용에 충분한 성능으로 판단하였음.



ResNet 분류 성능



Metrics	Measurement (Shape)	Measurement (Color)
Accuracy	0.9070	0.7907
Precision	0.8875	0.7982
Recall	0.9070	0.7907
F1-Score	0.8902	0.7863

향후 개선 필요 영역

- Oval, Other, Blue, Orange 등 소수 클래스 성능 부족

개선 방향

- 데이터 재수집/오버샘플링 등으로 클래스 불균형 완화
- 색·조명 변화에 대한 강화된 증강 적용
- 필요 시 클래스 병합·재정의를 통해 경계 명확화

분할선

Metrics	Values
Total number of test samples	150
Number of accurately detected samples	116
Overall accuracy	77.3%

향후 개선 필요 영역

- 분할선(Score Line) 검출 모듈은 정상적으로 동작하지만, 실제 데이터에서 분할선을 가진 약제가 매우 제한적
- 따라서 분할선 정보는 있으면 강력하지만, 전체 약제 중 일부에만 적용 가능한 보조 특징에 그친다는 한계 존재

개선 방향

- 분할선이 있는 약제 샘플을 추가 확보하여 학습·평가 데이터 다양성 확장
- 분할선이 없는 약제에도 활용 가능한 다른 구조적 특징(엠보싱, 캡슐 경계 등)을 함께 탐색하여 보조 특징군 확대

OCR 평가

기존 EasyOCR만 사용한 경우

약 번호	검출 신뢰도	모양 (신뢰도)	색상 (HSV)	색상 (CNN, 신뢰도)	최종 색상	분할선	각인 (신뢰도)
약 #1	86.28%	Circular (원형) (46.51%)	흰색, 블루	흰색 (28.66%)	흰색, 블루	없음	G R (53.30%)
약 #2	82.61%	Elliptical (타원형) (76.96%)	브라운, 핑크	브라운 (34.75%)	브라운, 핑크	없음	O (89.46%)
약 #3	78.86%	Elliptical (타원형) (74.49%)	흰색, 블루	핑크 (37.77%)	흰색, 블루	없음	없음
약 #4	78.67%	Oblong (장방형) (66.71%)	브라운, 블루	흰색 (30.59%)	브라운, 블루	없음	OI (26.62%)

Fallback으로 Google API를 사용한 경우

약 번호	검출 신뢰도	모양 (신뢰도)	색상 (HSV)	색상 (CNN, 신뢰도)	최종 색상	분할선	각인 (신뢰도)
약 #1	86.28%	Circular (원형) (46.56%)	흰색, 블루	흰색 (28.61%)	흰색, 블루	없음	GR (100.00%)
약 #2	82.59%	Elliptical (타원형) (76.99%)	브라운, 핑크	브라운 (34.84%)	브라운, 핑크	없음	BF (100.00%)
약 #3	78.87%	Elliptical (타원형) (74.54%)	흰색, 블루	핑크 (37.72%)	흰색, 블루	없음	M (100.00%)
약 #4	78.66%	Oblong (장방형) (66.66%)	브라운, 블루	흰색 (30.65%)	브라운, 블루	없음	HANWHA MUC100 (100.00%)

OCR 평가

향후 개선 필요 영역

- 기업 로고·특수 각인 오인식
 - 제약사 로고나 장식적인 심벌이 있는 경우, 이를 텍스트로 잘못 인식해 OCR 결과가 실제 각인과 불일치하는 경우가 발생함.
- 신뢰도(confidence)와 실제 정답의 괴리
 - 모델의 confidence 점수는 높게 나오지만, 사람이 확인했을 때는 잘못 인식된 사례가 일부 존재함.
- 각인 부재 면에 대한 처리 한계
 - 알약의 촬영 각도/방향에 따라 실제로 각인이 없는 면만 보이는 경우가 있어, 각인 기반 식별이 불안정해지는 구간이 존재함.

개선 방향

- 기업 로고를 독립적인 분류 기준으로 편입
 - 제약사 로고·심벌을 별도의 카테고리 정의하고, 로고 패턴에 대한 학습을 추가하여 로고/텍스트 오인식을 줄이고 전체 정확도를 향상.
- 각인 레이블 기준 정제 및 통일
 - 동일한 약제임에도 원본 데이터에서 각인 레이블 표기가 제각각인 경우가 있어, 이를 일관된 규칙으로 정규화하여 학습 데이터 품질을 향상.
- 휴먼 검수(Human-in-the-loop) 과정 도입
 - 고신뢰도이지만 내용이 의심되는 샘플, 또는 소수 클래스 샘플에 대해 전문가/사람이 직접 라벨을 검증·수정하는 검수 프로세스를 도입해, 점진적으로 ground truth 품질을 개선.

최종 Top-k 정확도 및 한계 분석

Top-9 정확도 : 55.06%

- 1,000장 테스트 이미지에서 추출된 4,170개 알약을 평가한 결과
- 전체 경구 약제 중 40% 이상(1,022종)이 특징이 거의 없는 흰색 원형 알약
 - 각인이 마모되거나 조명 반사, 뒤집지 않는 이상 각인 및 특징이 보이지 않는 경우가 많아 사람에게도 사실상 구분이 어려운 구조
- 따라서 단일 정답을 강제하기보다 Top-9 후보군을 제시하고 사용자가 최종 확인하도록 한 설계로 실제 현장에서 더 현실적이고 안전한 접근

논의 · 향후 계획 · 결론

향후 연구 방향

- 가장 큰 한계: 조명 반사·곡률로 인한 각인(OCR) 및 분할선 인식 실패
- Photometric Stereo 도입 검토
 - 여러 각도에서 찍은 이미지를 이용해 알약 표면의 3D 굴곡(음각·양각) 복원
 - 색상·반사 조건과 무관하게 각인 정보를 안정적으로 추출하는 것이 목표
- 3D 렌더링으로 다양한 각인·조명·각도 조합의 합성 데이터 생성, 제약 도메인 특화 OCR 모델 추가 학습(fine-tuning) 계획

의의 요약

- 단순 이미지 분류를 넘어 객체 탐지(다중 알약 검출) + 딥러닝 기반 형상·색상 특징 추출 + 각인 OCR + 분할선 분석을 하나의 파이프라인으로 통합한 실용적인 약제 식별 시스템 구현
- 복잡한 배경·혼합 약제 환경에서도 동작하는 실생활 지향형 AI 파이프라인이라는 점에서 의미
- 향후 스마트 약통, 헬스케어 디바이스, 조제 로봇 등에 탑재되는 약제 식별·복용 관리 자동화 모듈로의 가능성 보유

사회적 기대 효과

- 약사 업무의 계수·식별 부담 경감, 조제·투약 오류 예방
- 독거노인·시각장애인 등 스스로 약을 확인하기 어려운 분들에게 내가 먹는 약이 무엇인지를 알려주는 사회적 안전망 도구로 확장 가능

Q & A