

2023254019 이준혁

## <가짜리 문제풀이>

1. 딥러닝 비전 기술인 image classification, object detection, semantic/instance segmentation에 대해 설명하시오.

- Image classification: 객체의 종류를 class, label 등으로 구분하는 것으로 주어진 이미지를 사전 정의된 클래스 또는 카테고리 등으로 분류하는 것.
- Object detection: 여러 개의 객체에 대한 Localization + classification을 시각적으로 표시하는 것으로 경계상자로 그려서 물체의 존재 및 위치를 식별하는 것.
- Semantic Segmentation: 입력 영상을 픽셀 단위로 분류하는 것 → polygon 표시
- Instance segmentation: 시멘틱 분할 + 객체를 구분하여 폴리곤으로 표시함.  
(다양한 오브젝트를 표시함)

2. 2-stage detection의 특징과 R-CNN, FAST R-CNN, FASTER R-CNN을 비교 설명하시오

- 2stage detector: R-CNN, Fast R-CNN, Faster R-CNN 등 포함함  
물체 감지를 위해 2단계 네트워크를 사용하며 물체의 존재와 위치를 찾는 과정을 분리하여 수행
- R-CNN: 물체 감지를 위해 region proposal을 생성하고 각 region에 대하여 CNN을 적용하여 특징을 추출한 후 분류를 수행.  
높은 정확도를 보여주지만 속도가 느리고 region proposal 생성시 시간이 오래 걸림
- Fast-R-CNN: R-CNN의 속도나 정확도 개선을 위하여 개발됨.  
SVM을 softmax 함수로 변경 → Multi-task loss 함수로 분류와 위치를 동시에 최적화.  
동일한 region proposal을 이용하여 이따기만 한번만 CNN 연산 수행.
- Faster R-CNN: RPN: GPU수행.  
프리맵 픽셀당 9개의 Anchor box 적용.  
입력 이미지에 대해 sliding window를 적용하여 각 위치에서 물체의 존재 여부와 바운딩 박스를 예측함.

### 3. 다음의 객체검출 관련 용어를 설명 하시오.

#### (1) IOU : 위치 정확도 평가 지표

검출상자와 실물값과 여백값이 얼마나 겹치는지에 따른다.

$IOU > 0.5$  이면 제대로 검출한 것으로 인정함. 3/4의 경우 0.391 정도의 수치로 정확도가 높다고 할 수 없음.

#### (2) NMS : 비로대 억제.

- object detection 알고리즘은 객체가 있을 만한 위치에 수많은 Bbox를 출력
- 클러스터링으로 검출된 객체의 bounding box 중에 confidence가 가장 높은 box만 남기고 비슷한 위치에 있는 box는 중복성으로 간주하여 제거하는 기법.
- 중복여부는 box간  $IOU > 0.65$  이면 중복을 판단함.
- 미립된 경우는 good 검출을 제거하는 오류가 발생함.

#### (3) AP & mAP

AP: 모델의 성능을 하나의 값으로 정량적으로 비교하기 위한 평가 지표.

- 11-point 보간법  $\rightarrow$  2010 이후 all point.
- PR 곡선의 면적

mAP: 클래스별 AP를 평균한 값.

$$= \frac{1}{n} \sum_{i=1}^n AP_i$$

- mAP @ 0.5:  $IOU = 0.5$  (PASCAL VOC)

#### (4) Anchor box

미리 정해진 검출상자

- 대부분의 물체는 비슷한 모양을 가짐
- 학습 데이터의 Bbox를 대상으로 k-means clustering을 수행.
- $\rightarrow$  가장 대표적인 5개의 앵커 box 설정.
- Grid당 앵커수야량의 객체는 검출됨.

#### 4. 1Stage detector 인 YOLO가

다음 조건에서 객체검출을 수행한 결과로 출력하는 구조 설명

- 3x3 Grid: 이미지를 3x3으로 나누어 검출
- 2 Anchors: 각 그리드 셀당 2개의 앵커 사용
- 3 Classes: 자동차, 트럭, 사람 클래스.

$3 \times 3 \times [P_c, (b_x, b_y, b_h, b_w)]$

$(C_1, C_2, C_3) \times 2$

-  $P_c$ 는 objectionness score.

-  $b_x, b_y$ : 오프셋

-  $b_h, b_w$ : scaling factor

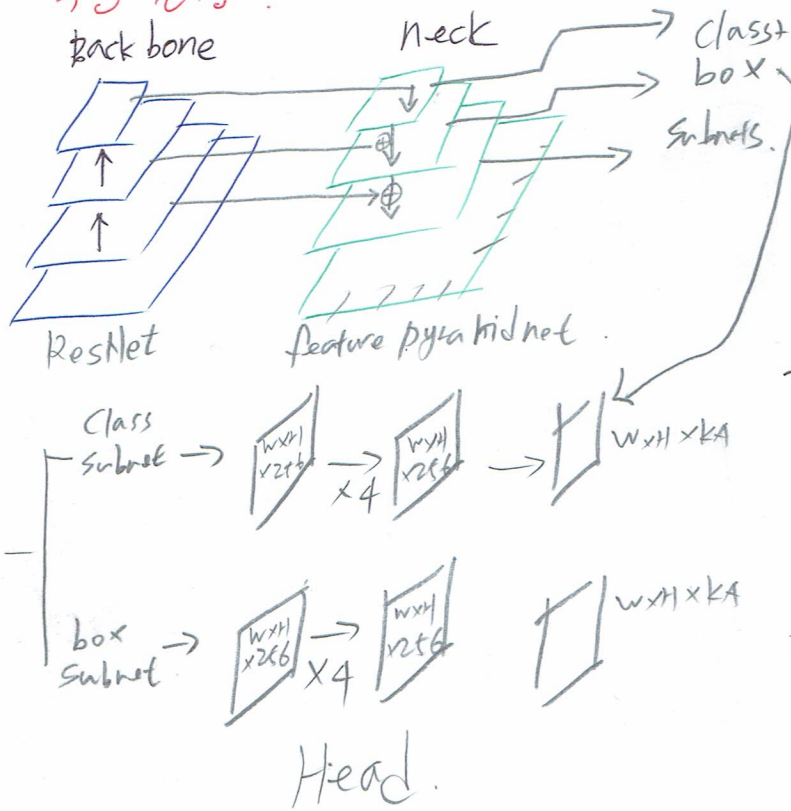
-  $C_1, C_2, C_3$ : confidence.

각 그리드 셀당 두개의 앵커 사용.

차원이 2가 붙음.



5. YOLO V3의 아키텍처를 Backbone, neck, head로 구분하여 그림 그 역할과 특징 설명.



• Backbone (Base): 입력 영상에서 특징맵을 추출함 (feature extractor)

• Neck: Backbone에서 생성된 특징을 융합 (mix + combine)

• Head: 각 bounding box에 대해 검출, classification + localization 수행.

6. YOLO V3의 FPN과 YOLO V4의 PAN에 대해서 설명하라.

- (V3) FPN (Feature pyramid network) →  
 (V4) PAN (Path Aggregation network)
- 작은 객체를 보다 잘 검출하기 위해서 다양한 크기의 피치맵을 이용.
  - (a) 하위 level의 피치맵 상시 정보와 상위 high level 피치맵의 융합된 정보를 융합.
  - (b) 융합의 효율성을 위해 bottom-up path를 추가 (red long path → blue 'shortcut' path)
  - (c) (V3) scale 별 검출 수행 → (V4) 모든 피치맵을 융합 / 융합한 후 검출 수행
- (a) FPN backbone  
 (b) bottom-up path argumentation.  
 (c) adaptive feature pooling.

## 17. 데이터를 가져와서 학습 행위에 대해 설명하라.

- PASCAL VOC (Visual object classes)
  - XML format.
  - 20개 오브젝트 카테고리.
  - AP에서 표준이었는데 최근에 잊혀짐.
- MS COCO
  - 객체검출 성능평가의 표준이 되는 데이터셋.
- Google open images.
  - 클래스의 개수무수.
- Kaggle.
  - 데이터 탐색과 분석 방법, 코드.
- AI HUB.
  - 양질의 대용량 인공지능 학습용 데이터 보유.
- 인공지능 제로플랫폼 (KAMP)
  - 제로 AI 데이터셋 + 분석 학습 가이드북.
- 지식/레이블.
  - 영상이란 위치 클래스, 속성 등의 정보를 갖고 있는 텍스트로 된 설명타입
  - 지식 타입의 행아웃 주로 xml, json, csv, txt를 사용함.

## • YOLOv5의 학습 행위에 대해.

- Class ID: 0~79 for COCO Dataset.
- Bounding box:  $ncx, ncy, nw, nh$   
[0 ~ 1]