

# 9강 질문 정답

## 1. mAP는 어떻게 계산될 수 있는지 설명

mAP (mean Average Precision)는 객체 검출 모델의 성능을 평가하는 주요 지표 중 하나입니다. mAP는 여러 클래스에 대한 Average Precision(AP)를 평균하여 계산됩니다.

- **Average Precision (AP):**
  1. **정확도(Precision)와 재현율(Recall) 계산:** 모델이 예측한 바운딩 박스들을 신뢰도(score) 순으로 정렬합니다. 각 박스에 대해 True Positive(TP)와 False Positive(FP)를 결정하여 Precision과 Recall 값을 계산합니다.
  2. **Precision-Recall 곡선 작성:** 다양한 임계값에서 Precision과 Recall 값을 플롯하여 곡선을 그립니다.
  3. **AP 계산:** Precision-Recall 곡선 아래의 면적을 계산하여 AP를 도출합니다.
- **mAP 계산:** 여러 클래스에 대해 AP를 계산한 후, 이를 평균하여 mAP를 구합니다. 이는 모델의 전반적인 성능을 종합적으로 평가하는 데 사용됩니다.

---

## 2. SPP(RoI Pooling)에 대해 설명

SPP (Spatial Pyramid Pooling)는 RoI Pooling의 개선된 버전으로, 다양한 크기의 입력 이미지를 고정된 크기의 특징 맵으로 변환하는 기술입니다.

- **RoI Pooling:**
  - 서로 다른 크기의 RoI(Region of Interest)를 고정된 크기의 그리드로 분할하고, 각 그리드에서 최대값을 추출하여 고정된 크기의 출력 벡터를 생성합니다.
- **SPP의 장점:**
  - **다중 스케일 처리:** 여러 크기의 풀링 창을 사용하여 다양한 스케일의 특징을 동시에 추출합니다.
  - **유연성 향상:** 입력 이미지의 크기에 관계없이 고정된 크기의 출력 벡터를 생성할 수 있어, 네트워크의 후속 레이어에 효율적으로 전달됩니다.

SPP는 특히 다양한 크기의 객체를 효과적으로 검출하는 데 유리합니다.

---

### 3. RoI Projection에 대해 설명

RoI Projection은 Region of Interest(RoI)를 네트워크의 다른 단계나 다른 공간으로 변환하는 과정을 의미합니다.

- **주요 기능:**
  - **다중 스케일 매핑:** 예를 들어, Feature Pyramid Network(FPN)에서는 여러 스케일의 특징 맵이 존재합니다. RoI Projection을 통해 특정 RoI를 가장 적합한 스케일의 특징 맵에 매핑합니다.
  - **정확도 향상:** 적절한 스케일의 특징 맵에 RoI를 투영함으로써, 객체의 위치와 크기에 대한 더 정확한 예측이 가능합니다.
- **적용 예시:**
  - **FPN 구조에서의 활용:** 각 RoI가 어떤 스케일의 특징 맵에 속하는지 결정하여, 해당 스케일의 특징을 활용해 더 정밀한 검출을 수행합니다.

RoI Projection은 다양한 스케일의 정보를 효과적으로 활용하여 객체 검출의 정확도를 높이는 데 중요한 역할을 합니다.

---

### 4. Faster RCNN 중 RPN에 대해 설명

RPN (Region Proposal Network)는 Faster R-CNN의 핵심 구성 요소로, 이미지 내에서 객체가 존재할 가능성이 높은 영역을 제안하는 역할을 합니다.

- **작동 방식:**
  1. **슬라이딩 윈도우:** 전체 이미지의 특징 맵에서 슬라이딩 윈도우 방식으로 이동하면서 각 위치에서 여러 앵커 박스를 평가합니다.
  2. **앵커 박스 평가:** 각 앵커 박스에 대해 객체 존재 확률과 바운딩 박스 조정을 예측합니다.
  3. **지역 제안:** 높은 확률을 가진 앵커 박스들을 선택하여 RoI(Region of Interest)로 제안합니다.
- **장점:**
  - **속도 향상:** 별도의 지역 제안 알고리즘 없이도 빠르게 지역을 제안할 수 있습니다.
  - **통합 학습:** RPN과 객체 검출 네트워크가 동시에 학습되어 전체적인 성능이 향상됩니다.

RPN은 Faster R-CNN의 전체적인 객체 검출 성능과 속도를 크게 개선하는 중요한 역할을 합니다.

---

## 5. Anchor box에 대해 설명

Anchor box는 객체 검출에서 다양한 크기와 비율의 사전 정의된 바운딩 박스입니다. 이는 네트워크가 다양한 형태와 크기의 객체를 효과적으로 검출할 수 있도록 돕습니다.

- **특징:**
  - **다양한 크기와 비율:** 다양한 크기와 비율을 가진 여러 앵커 박스를 사용하여, 다양한 형태의 객체에 대응합니다.
  - **격자 기반 배치:** 이미지의 각 위치에 여러 앵커 박스를 배치하여, 모든 가능한 객체 위치를 커버합니다.
- **활용:**
  - **위치 및 크기 조정:** 네트워크는 각 앵커 박스에 대해 객체의 존재 여부를 예측하고, 위치와 크기를 조정하여 최종 바운딩 박스를 생성합니다.
  - **학습의 안정성:** 다양한 앵커 박스를 사용함으로써, 다양한 객체의 위치와 크기에 대한 정보를 네트워크가 학습할 수 있습니다.

Anchor box는 특히 다중 객체 검출에서 중요한 역할을 하며, 객체의 다양성을 효과적으로 처리할 수 있게 합니다.

---

## 6. RPN의 역할에 대해 설명

RPN (Region Proposal Network)의 주요 역할은 이미지 내에서 객체가 존재할 가능성이 높은 지역을 빠르고 효율적으로 제안하는 것입니다.

- **세부 역할:**
  1. **지역 제안:** 슬라이딩 윈도우 방식으로 전체 이미지의 특징 맵을 탐색하면서, 각 위치에서 여러 앵커 박스를 평가하여 객체 존재 확률과 바운딩 박스 조정을 예측합니다.
  2. **고품질 후보 생성:** 높은 확률을 가진 앵커 박스들을 선택하여 RoI(Region of Interest)로 제안함으로써, 후속 단계에서 처리할 후보 지역을 제한합니다.
  3. **통합 학습:** RPN은 객체 검출 네트워크와 동시에 학습되어, 전체적인 모델의 성능을 향상시킵니다.
- **장점:**
  - **속도 향상:** 별도의 지역 제안 알고리즘 없이도 빠르게 후보 지역을 생성할 수 있습니다.
  - **효율성:** 제한된 수의 후보 지역만을 후속 단계에서 처리함으로써, 전체 연산량을 줄이고 효율성을 높입니다.

RPN은 Faster R-CNN과 같은 2-stage 객체 검출 모델에서 핵심적인 역할을 하며, 전체 검출 성능과 속도에 큰 영향을 미칩니다.

---

## 7. Neck의 역할에 대해 설명

Neck은 객체 검출 네트워크에서 백본(backbone)과 헤드(head) 사이에 위치한 구성 요소로, 다중 스케일의 특징을 효과적으로 결합하고 전달하는 역할을 합니다.

- **주요 기능:**
  - **특징 통합:** 백본에서 추출한 다양한 스케일의 특징 맵을 결합하여 풍부한 정보를 생성합니다.
  - **다중 스케일 처리:** 작은 객체부터 큰 객체까지 다양한 크기의 객체를 효과적으로 검출할 수 있도록 다중 스케일의 특징을 활용합니다.
  - **정보 전달:** 통합된 특징을 헤드 단계로 전달하여, 최종적으로 객체의 위치와 클래스를 예측하는 데 사용됩니다.
- **대표적인 구조:**
  - **FPN (Feature Pyramid Network):** 여러 스케일의 특징 맵을 피라미드 형태로 결합하여 다중 스케일 정보를 제공합니다.
  - **PANet (Path Aggregation Network):** FPN을 개선하여 정보 흐름을 강화하고, 상향 경로를 추가하여 더 풍부한 특징을 생성합니다.

Neck은 전체 네트워크의 성능을 향상시키는 데 중요한 역할을 하며, 특히 다양한 크기의 객체를 효과적으로 검출하는 데 기여합니다.

---

## 8. FPN, PANet 에 대해 설명

FPN (Feature Pyramid Network)와 PANet (Path Aggregation Network)은 객체 검출에서 다중 스케일의 특징을 효과적으로 활용하기 위해 설계된 네트워크 구조입니다.

### FPN (Feature Pyramid Network)

- **개요:** FPN은 다양한 스케일의 특징 맵을 피라미드 형태로 결합하여, 각 스케일에서 풍부한 정보를 제공합니다.
- **구조:**

- **Bottom-Up Pathway:** 전통적인 백본 네트워크(예: ResNet)의 여러 레이어에서 특징 맵을 추출합니다.
- **Top-Down Pathway:** 고해상도의 특징 맵을 생성하기 위해, 상위 레이어의 특징을 업샘플링하여 하위 레이어의 특징과 결합합니다.
- **측면 연결(Side Connections):** 서로 다른 스케일의 특징을 결합하여 다중 스케일의 정보를 통합합니다.
- **장점:**
  - **다중 스케일 처리:** 작은 객체부터 큰 객체까지 다양한 크기의 객체를 효과적으로 검출할 수 있습니다.
  - **효율성:** 추가적인 계산 비용 없이 다중 스케일의 정보를 활용할 수 있습니다.

## PANet (Path Aggregation Network)

- **개요:** PANet은 FPN을 개선한 구조로, feature map의 정보 흐름을 강화하기 위해 상향 경로를 추가한 네트워크입니다.
- **구조:**
  - **Bottom-Up Path Enhancement:** 상향 경로를 통해 하위 레이어의 정보를 상위 레이어로 전달하여, 특징 맵 간의 정보 교환을 강화합니다.
  - **경로 연결:** 여러 경로를 통해 특징 맵의 정보를 통합하여, 더 풍부한 표현력을 제공합니다.
- **장점:**
  - **정보 흐름 강화:** FPN보다 더 많은 정보가 특징 맵 간에 전달되어, 검출 성능이 향상됩니다.
  - **소형 객체 검출 개선:** 상향 경로를 통해 소형 객체의 특징을 더욱 잘 포착할 수 있습니다.

FPN과 PANet은 모두 다중 스케일의 특징을 효과적으로 활용하여 객체 검출의 정확도와 효율성을 높이는 데 기여합니다.

## 9. 2 stage와는 다르게 RPN이 없는 1 stage에서 어떻게 박스를 예측하는지 설명

1-stage 객체 검출기는 RPN과 같은 별도의 지역 제안 네트워크 없이, 직접적으로 바운딩 박스와 클래스 확률을 예측합니다. 대표적인 1-stage 모델로는 YOLO, SSD 등이 있습니다.

- **작동 방식:**
  1. **격자 기반 예측:** 입력 이미지를  $S \times S$  격자로 나누고, 각 격자 셀에서 객체의 존재 여부를 예측합니다.

2. **앵커 박스 활용:** 각 격자 셀에서 여러 앵커 박스에 대해 객체의 클래스 확률과 바운딩 박스의 위치 및 크기를 예측합니다.
3. **동시 예측:** 클래스 확률과 바운딩 박스를 동시에 예측하여, 객체를 검출합니다.

- **장점:**

- **속도:** 단일 네트워크 패스로 모든 예측을 수행하기 때문에, 2-stage 모델보다 훨씬 빠릅니다.
- **단순성:** 구조가 단순하여 구현과 학습이 용이합니다.

- **단점:**

- **정확도:** RPN을 사용하는 2-stage 모델에 비해 정확도 면에서 다소 낮을 수 있습니다.
- **소형 객체 검출:** 작은 객체 검출에 상대적으로 약할 수 있습니다.

1-stage 모델은 주로 실시간 객체 검출이 필요한 응용 분야에서 많이 사용되며, 속도와 효율성이 중요한 경우에 적합합니다.

---

## 10. Yolo v1에 대해 설명

YOLO v1 (You Only Look Once version 1)는 단일 신경망을 사용하여 이미지 내의 객체를 빠르게 검출하는 1-stage 객체 검출 모델입니다.

- **작동 방식:**

1. **그리드 분할:** 입력 이미지를  $S \times S$  격자로 나눕니다.
2. **바운딩 박스 예측:** 각 격자 셀에서 B개의 바운딩 박스를 예측합니다. 각 박스는 위치 좌표, 너비, 높이, 신뢰도(confidence)로 구성됩니다.
3. **클래스 확률 예측:** 각 격자 셀에서 C개의 클래스에 대한 확률을 예측합니다.
4. **최종 예측:** 신뢰도와 클래스 확률을 결합하여 최종 객체를 검출합니다.

- **주요 특징:**

- **단일 네트워크 패스:** 전체 이미지를 한 번에 처리하여 빠른 속도를 구현합니다.
- **전역 컨텍스트 활용:** 이미지 전체의 컨텍스트를 고려하여 예측함으로써, 예측의 일관성과 정확성을 높입니다.
- **End-to-End 학습:** 전체 검출 과정을 통합적으로 학습하여, 최적화가 용이합니다.

- **장점:**

- **속도:** 매우 빠른 객체 검출이 가능하여 실시간 애플리케이션에 적합합니다.
- **단순성:** 구조가 단순하여 구현이 용이합니다.

- **단점:**

- **정확도 제한:** 작은 객체나 밀집된 객체 검출에서 성능이 다소 낮을 수 있습니다.

- **고정된 앵커 박스:** 다양한 비율의 객체를 잘 처리하지 못할 수 있습니다.

YOLO v1은 객체 검출의 속도와 효율성을 크게 향상시켰으나, 이후 버전에서 정확도와 다양한 기능이 추가되었습니다.

## 11. EfficientDet의 등장 배경 및 Compound scaling에 대해 설명

EfficientDet은 객체 검출에서 높은 정확도와 효율성을 동시에 달성하기 위해 개발된 모델입니다. EfficientDet은 EfficientNet을 기반으로 하여, 백본(backbone), Neck, 헤드를 효율적으로 설계하였습니다.

### 등장 배경

- **효율성 필요성:** 기존의 객체 검출 모델은 정확도와 효율성 간의 trade-off가 존재하여, 둘 다 최적화하는 것이 어려웠습니다.
- **EfficientNet의 성공:** EfficientNet은 이미지 분류에서 높은 효율성과 정확도를 달성하였으며, 이를 객체 검출에 적용하고자 했습니다.
- **모델 통합 최적화:** 백본, Neck, 헤드 모두를 통합적으로 최적화하여, 전체 모델의 효율성을 극대화하고자 했습니다.

### Compound Scaling

Compound Scaling은 네트워크의 깊이(depth), 너비(width), 해상도(resolution)를 균형 있게 확장하는 방법입니다.

- **방법:**
  - **균형적 확장:** 깊이, 너비, 해상도를 단순히 개별적으로 확장하는 대신, 모든 차원을 동시에 고려하여 균형 있게 확장합니다.
  - **비율 조정:** 특정 비율로 각 차원을 확장하여, 전체 네트워크의 효율성과 성능을 최적화합니다.
- **장점:**
  - **효율성:** 단순한 확장 방식보다 더 효율적으로 네트워크를 확장할 수 있습니다.
  - **성능 향상:** 균형 잡힌 확장을 통해, 다양한 크기의 모델에서도 높은 성능을 유지할 수 있습니다.

EfficientDet은 Compound Scaling을 통해 다양한 크기의 모델을 제공하여, 다양한 애플리케이션의 요구에 맞게 선택할 수 있으며, 높은 정확도와 효율성을 동시에 달성하였습니다.

---

## 12. Cascade, Deformable, Swin 에 대해 설명

### Cascade

Cascade 객체 검출기는 여러 단계의 검출기를 순차적으로 적용하여, 점진적으로 검출 성능을 향상시키는 방식입니다.

- 작동 방식:
  1. 단계적 처리: 첫 번째 단계에서 기본적인 객체 검출을 수행하고, 이후 단계에서 더 높은 IoU(Intersection over Union) 임계값을 적용하여 검출 결과를 정교화합니다.
  2. 점진적 개선: 각 단계에서 이전 단계의 결과를 기반으로, 검출의 정확도를 높여 나갑니다.
- 장점:
  - 정확도 향상: 단계별로 정교한 검출을 수행하여, 최종 결과의 정확도를 높입니다.
  - 고품질 검출: 높은 IoU 임계값을 통해, 더 정확한 바운딩 박스를 생성합니다.
- 단점:
  - 연산 비용 증가: 여러 단계를 거치기 때문에, 계산 비용이 증가할 수 있습니다.

Cascade는 특히 높은 정확도가 요구되는 응용 분야에서 유용하게 사용됩니다.

### Deformable

Deformable Convolutional Networks (DCN)은 표준 합성곱 연산을 변형하여, 입력 특징 맵의 적응적인 위치에 필터를 적용하는 방식입니다.

- 특징:
  - 적응적 필터 위치: 각 합성곱 필터가 고정된 위치가 아닌, 입력 특징 맵의 데이터에 따라 변형된 위치에서 연산을 수행합니다.
  - 유연성 향상: 다양한 형태와 크기의 객체를 더 잘 처리할 수 있도록, 필터가 입력 데이터에 맞게 유연하게 조정됩니다.
- 장점:
  - 비정형 객체 처리: 비정형적이거나 다양한 형태의 객체를 효과적으로 검출할 수 있습니다.
  - 성능 향상: 표준 합성곱보다 더 높은 표현력을 가지며, 객체 검출 정확도를 향상시킵니다.
- 적용 예시:



- **Deformable Faster R-CNN:** Deformable Convolution을 적용하여 Faster R-CNN의 성능을 개선한 모델입니다.

Deformable 연산은 특히 복잡한 형태의 객체를 검출하는 데 유용하게 사용됩니다.

## Swin

**Swin Transformer**는 이미지 처리에 Transformer 구조를 적용한 모델로, Shifted Windowing을 통해 효율적인 연산과 높은 표현력을 동시에 달성합니다.

- **특징:**
  - **Shifted Windowing:** 이미지를 윈도우 단위로 분할하여 self-attention을 수행하고, 다음 계층에서는 윈도우를 이동시켜 교차 윈도우 간의 정보 교환을 촉진합니다.
  - **계층적 구조:** 여러 계층을 쌓아올리면서, 다양한 스케일의 특징을 추출합니다.
  - **효율성:** 전역 self-attention 대신 지역적 self-attention을 수행하여, 연산 비용을 줄이고 효율성을 높입니다.
- **장점:**
  - **높은 성능:** 객체 검출을 포함한 다양한 비전 작업에서 높은 성능을 보입니다.
  - **유연성:** 다양한 크기의 모델로 확장 가능하며, 다양한 애플리케이션에 적용할 수 있습니다.
- **적용 예시:**
  - **Swin Transformer 기반 객체 검출기:** Swin을 백본으로 사용하는 객체 검출 모델들은 높은 정확도와 효율성을 동시에 달성합니다.

Swin Transformer는 Transformer의 강력한 표현력을 이미지 처리에 효과적으로 적용한 모델로, 최근 객체 검출 분야에서 많이 사용되고 있습니다.

---

## 13. M2Det에 대해 설명

**M2Det (Multi-level Feature Pyramid Network for Object Detection)**은 다중 스케일의 특징을 효과적으로 활용하여 객체 검출 성능을 향상시키는 1-stage 모델입니다.

- **주요 구성 요소:**
  1. **Multi-level Feature Pyramid Network (M2FPN):**
    - **다중 스케일 특징 통합:** 여러 레벨의 특징 맵을 결합하여 풍부한 다중 스케일 정보를 생성합니다.

- **다층 피라미드 구조:** 다양한 크기의 객체를 효과적으로 검출할 수 있도록, 여러 층의 특징 맵을 통합합니다.

## 2. Multi-level Prediction Module (MPM):

- **다중 스케일 예측:** 결합된 다중 스케일 특징을 활용하여, 각 레벨에서 바운딩 박스와 클래스 확률을 예측합니다.
- **효율적 예측:** 단일 단계에서 다양한 스케일의 객체를 정확하게 검출할 수 있도록 설계되었습니다.

### • 특징:

- **높은 정확도:** 다중 스케일의 특징을 효과적으로 활용하여, 다양한 크기의 객체를 정확하게 검출합니다.
- **효율성:** 1-stage 구조로 인해, 2-stage 모델에 비해 빠른 속도를 유지하면서도 높은 정확도를 달성합니다.
- **유연성:** 다양한 백본 네트워크와 쉽게 통합될 수 있어, 다양한 응용 분야에 적용할 수 있습니다.

### • 장점:

- **다중 스케일 처리:** 다양한 크기의 객체를 효과적으로 검출할 수 있어, 실제 응용에서 유용합니다.
- **단순한 구조:** 1-stage 구조로 구현이 단순하며, 실시간 객체 검출에 적합합니다.
- **높은 성능:** 여러 벤치마크에서 높은 정확도를 기록하여, 경쟁력 있는 모델로 평가받고 있습니다.

M2Det은 다중 스케일의 특징을 효과적으로 활용하여, 높은 정확도와 효율성을 동시에 달성하는 1-stage 객체 검출 모델로, 다양한 실용적인 응용 분야에서 유용하게 사용됩니다.