


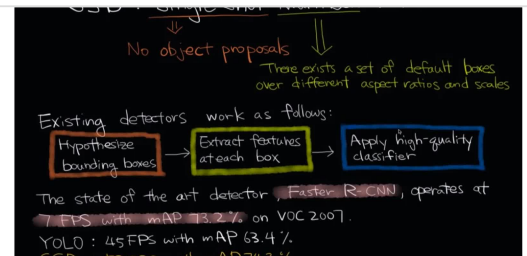
# Abstract

## Image Detection 방법론: AttentionNet, SSD, YOLO, YOLOv2

[LECTURE] Image Detection 방법론: AttentionNet, SSD, YOLO, YOLOv2 : edwith

학습목표 앞 강의에 이어 이번 강의에서도 Detection 방법론들을 알아보도록 하겠습니다. 이번 강의에서는 앞서 배운 R-CNN 계열의 방법들과는 다르게 새로운 방법들로, YOL... - 커넥트재단

 <https://www.edwith.org/deeplearningchoi/lecture/15579?isDesc=false>



## AttentionNet: Aggregating Weak Directions for Accurate Object Detection

bounding box 의 크기를 순차적으로 줄여서 보다 정확한 물체의 위치에 bounding box 를 맞추는 아이디어.

논문의 영향:

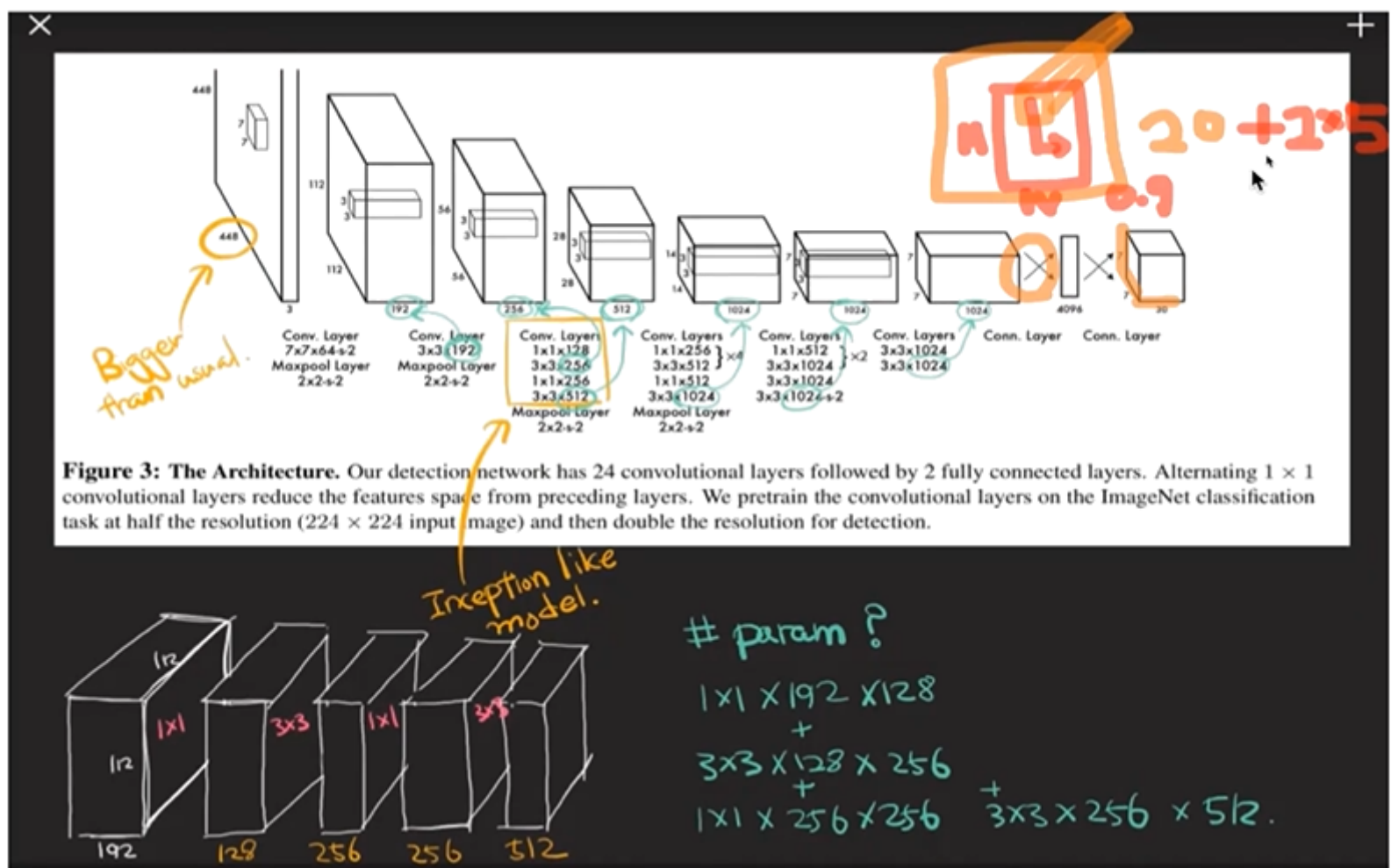
- Proposed a novel detection method, which estimates an exact bounding box by aggregating weak predictions.
- Does not sperate object proposal, classifiers, post bounding box regressions.
- State-of-the art on single class object detectin tasks.

## YOLO: You only look once

### YOLO Detection system

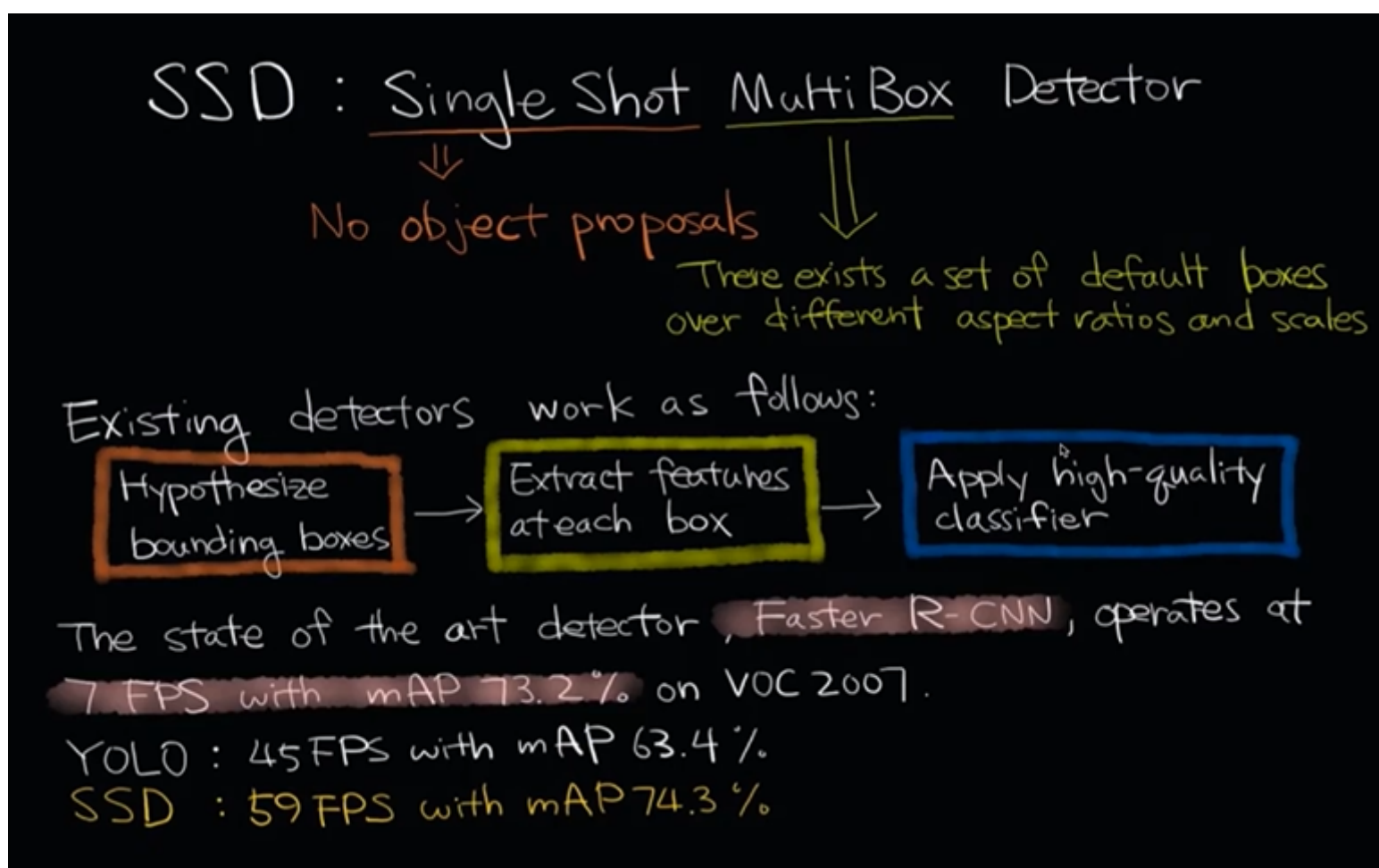
Simultaneously predicts multiple bounding boxes and class probabilities. ⇒ Frame detection problem as a regression problem.

### Scheme

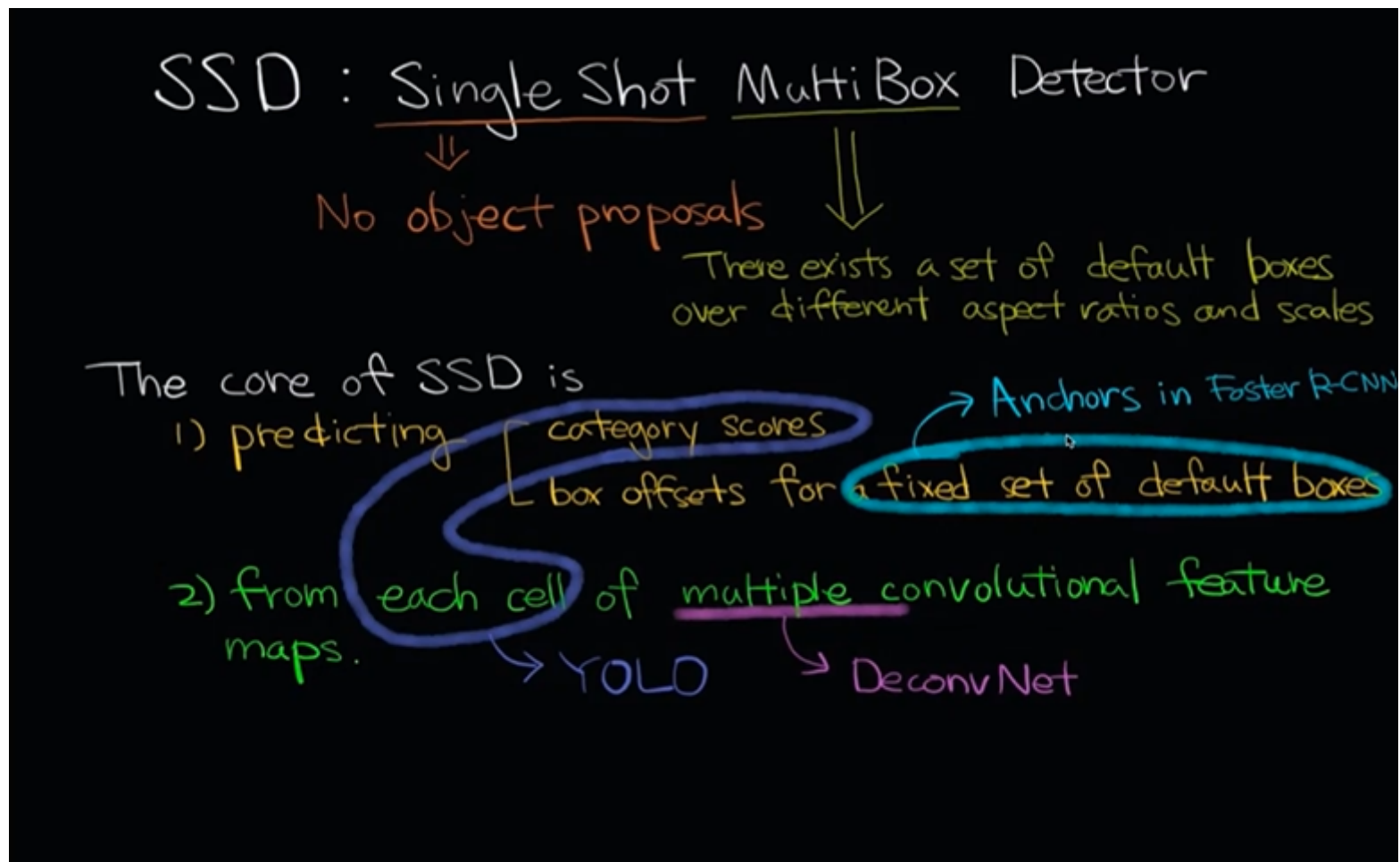


## SSD: Single Shot MultiBox Detector

YOLO 의 컨셉과 Faster RCNN 의 RPN 컨셉을 합친 것.



## Core of SSD



## YOLO 9000: Better, Faster, Stronger

YOLO 가 SSD 보다 성능이 잘 나오기 위해서 여러 실험을 통해 괜찮은 테크닉 노하우를 얻음.

**Original YOLO suffers from several shortcomings.**

1. Localization errors
2. Low recall compared region proposal based method.
  - Precision  
정밀도란 모델이 True 라고 분류한 것 중에서 실제 참인 것의 비율이다. 정답률과 동일한 의미.
  - Recall  
재현율이란 실제 참인 것중에서 모델이 참이라고 예측한 것의 비율이다. 정밀도와 true positive 의 경우를 다르게 바라보는 것이다. 즉, precision 과 recall 은 모두 실제 true 인 정답을 모델이 true 라고 예측한 경우에 관심이 있으나, 바라보고자 하는 관점만 다르다.
  - more specific

### Some strategies for 'Faster'

1. Batch Normalization
2. High Resolution : [224 224] → [448 448]
3. Anchor Boxes : Used anchor boxes similar to RPNs.
4. Dimension Clusters : Determine the size of anchor boxes with clustering.
5. Location prediction : Instead of predictin locations, YOLO 9000 predicts box offsets and scaling of box size.
6. Fine-grained Features

Original YOLO uses 13\*13 feature maps, YOLO 9000 incorporates previous 26\*26 feature map.

#### 7. Multi-scale training

Since YOLO 9000 uses conv and pooling, any arbitrary input dimension can be used.

### **Some strategies for 'Stronger'**

- Hierarchical classification