10 目标检测

任务、方法思路与演进

学习目标

- 说明目标检测的输入输出及主要挑战。
- 解释区域检测、单阶段检测与集合预测的基本思路。
- 说明代表方法改变了什么，以及改变的原因。
- 从候选组织、框表示和预测对应关系比较不同路线。
- 理解检测评价的含义与方法比较所需的条件。

1 检测任务与共同工具

从整图分类走向多实例类别与位置，理解方法要解决的问题

从整图类别到多个目标

- **目标检测**为每个关注的目标实例预测类别与边界框。
- 图中狗和猫各有一个框；同类实例也需分别检测，目标数可以变化或为零。

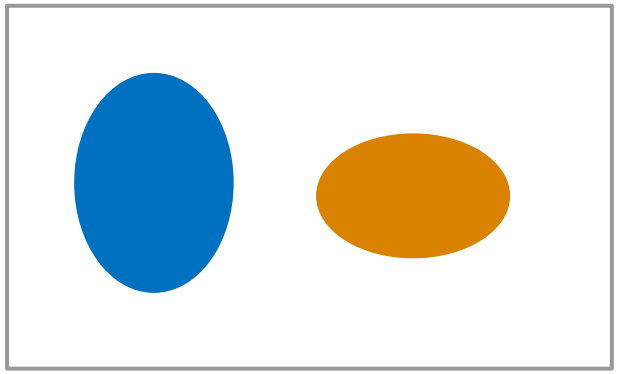


教学标框，不是模型推理结果

分类、检测与分割的输出

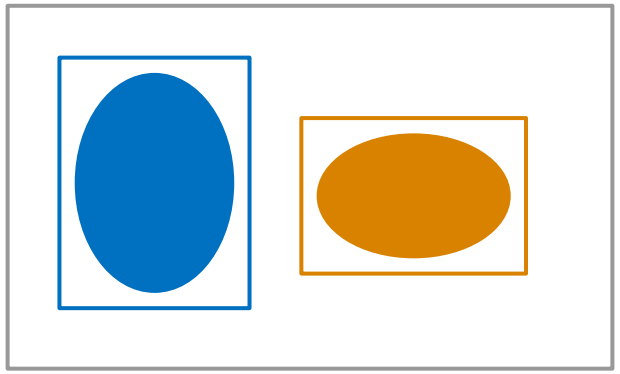
- 分类描述整图，检测区分实例并给出框，语义分割给每个像素赋类别。
- 下图使用相同的两个示意目标；本讲的轴对齐矩形框可能包含背景或遮挡区域。

原图 / 整图分类



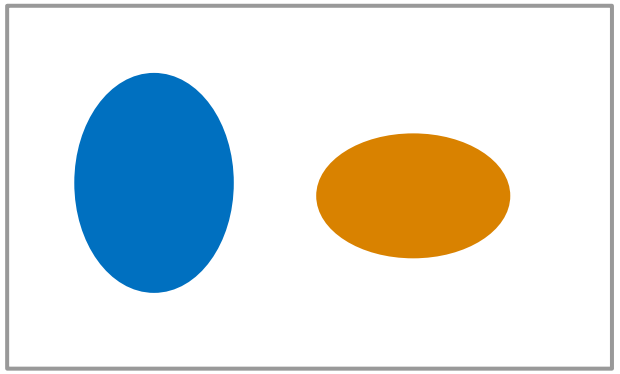
整图标签：含目标

检测：实例类别与框



框包含目标及部分背景

分割：逐像素类别



蓝 / 橙像素分别赋类别

从图像到一组目标预测

- 标注是一组“类别、边界框”，排列不影响含义；预测另附用于筛选和排序的分数。
- 框可用左上角、右下角坐标表示；训练使用标注，推理只输入图像。
- 先以区域方法为例，形成从图像到检测结果的整体参照。
- 后续比较预测来源、框表示和监督对应，观察各方法改变哪个环节。

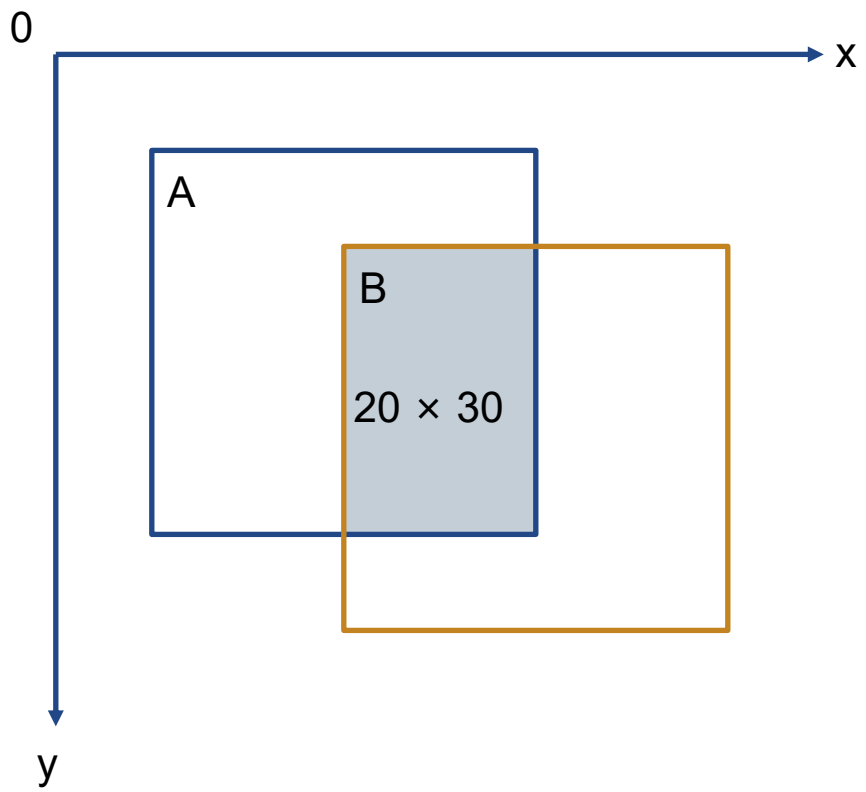


检测方法需要解决哪些困难

困难	对方法设计的要求
目标位置与数量变化	在图中寻找多个可能的目标
尺度与形状变化	用合适的空间信息描述大小不同的目标
背景多、目标少，且可能遮挡	区分类别、背景与相邻实例
待检查位置多	控制特征计算和候选处理的成本

这些要求共同影响候选组织、特征表示和预测方式。

IoU：比较两个框的重叠



A = (10, 10, 50, 50)

B = (30, 20, 70, 60)

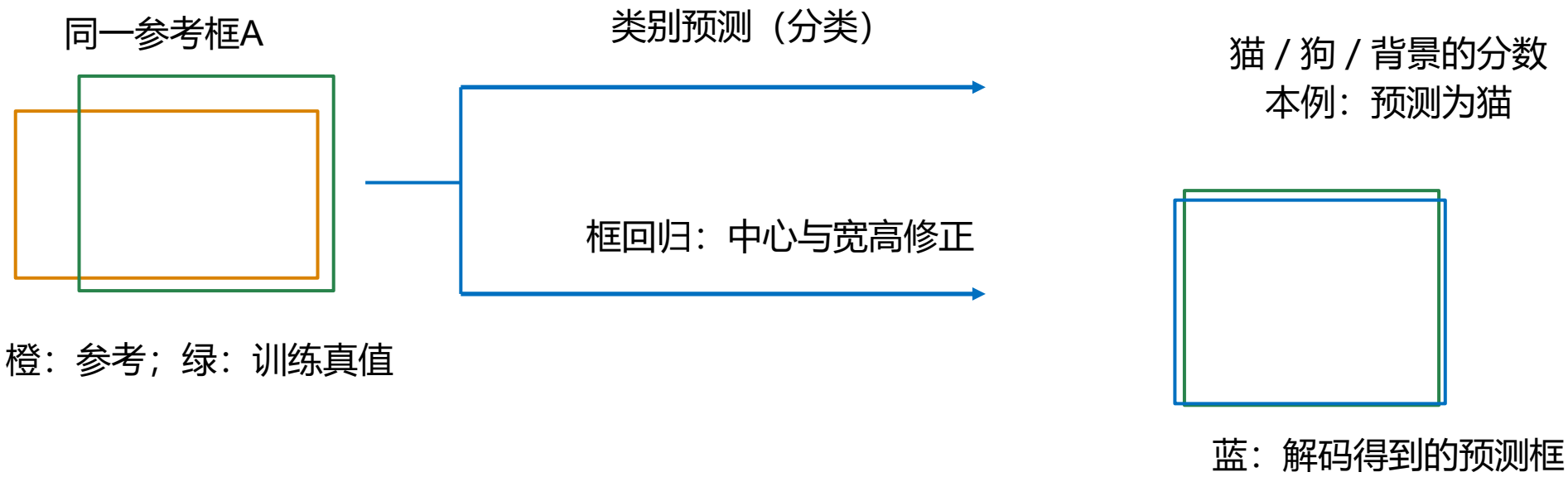
- **交并比（IoU）**为交集面积除以并集面积，取值0到1。
- 图中两框面积各1600、交集600，

$$\mathrm{IoU} = \frac{600}{1600+1600-600} \approx 0.231。$$

- 训练分配：参考框与真实框比较，确定哪些预测学习哪个目标。
- 结果筛选：比较预测框之间的IoU，识别高度重叠的重复结果。
- 检测评价：比较预测框与真实框，判断IoU是否达到定位阈值。

同时学习类别与位置

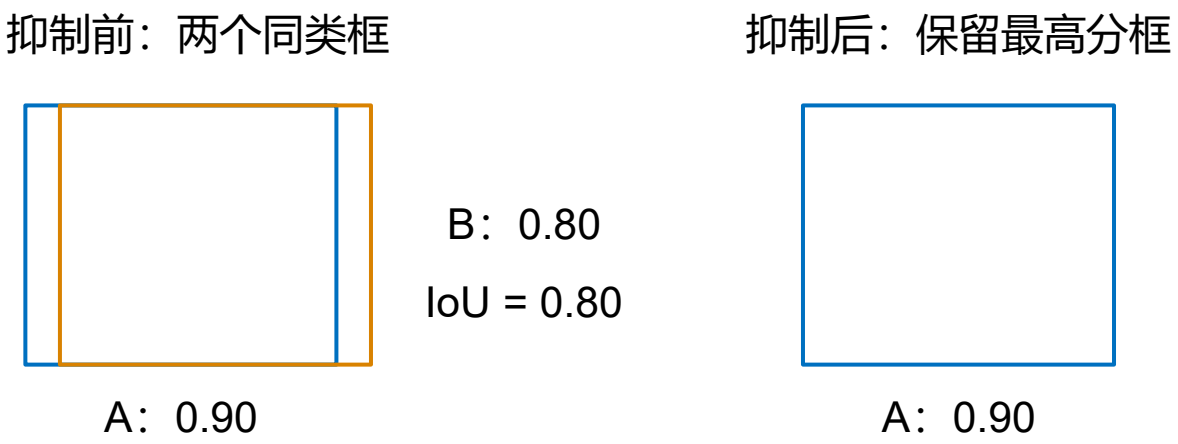
- 推理从图像产生类别与框，筛选后得到最终检测结果。
- 训练先将预测对应真实目标，再确定类别 / 位置目标，计算损失并反向传播优化。
- 参考框示例中，类别预测判断类别或背景，框回归学习位置修正；忽略样本不计对应损失。



训练：正样本学习类别与位置；负样本只学习背景类别

重复预测与非极大值抑制

- 同一目标可被多个预测共同学习；推理后可能出现多个描述同一只猫的高分框，需要筛除重复。
- **非极大值抑制（NMS）**在同类预测中保留当前最高分框，抑制与它高度重叠的框，再重复处理剩余框。
- 示例阈值为0.5，A、B的IoU为0.8，因此保留A、抑制B；拥挤时也可能误抑制相邻目标。



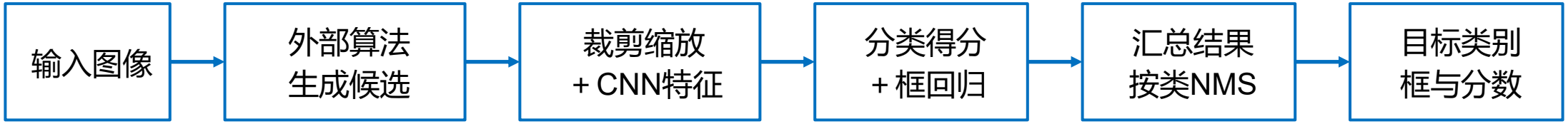
2 区域检测与共享计算

重复预测可以筛除，哪些区域值得识别？从候选搜索进入区域计算与共享

R-CNN：完整检测流程

- R-CNN把分类网络用于检测：输入整张图像，输出各目标的类别、框与分数。
- **候选区域**减少待检查区域，仍可能是背景；R-CNN对每个候选分别运行CNN。

推理：使用训练好的模型，从一张图像得到多个目标



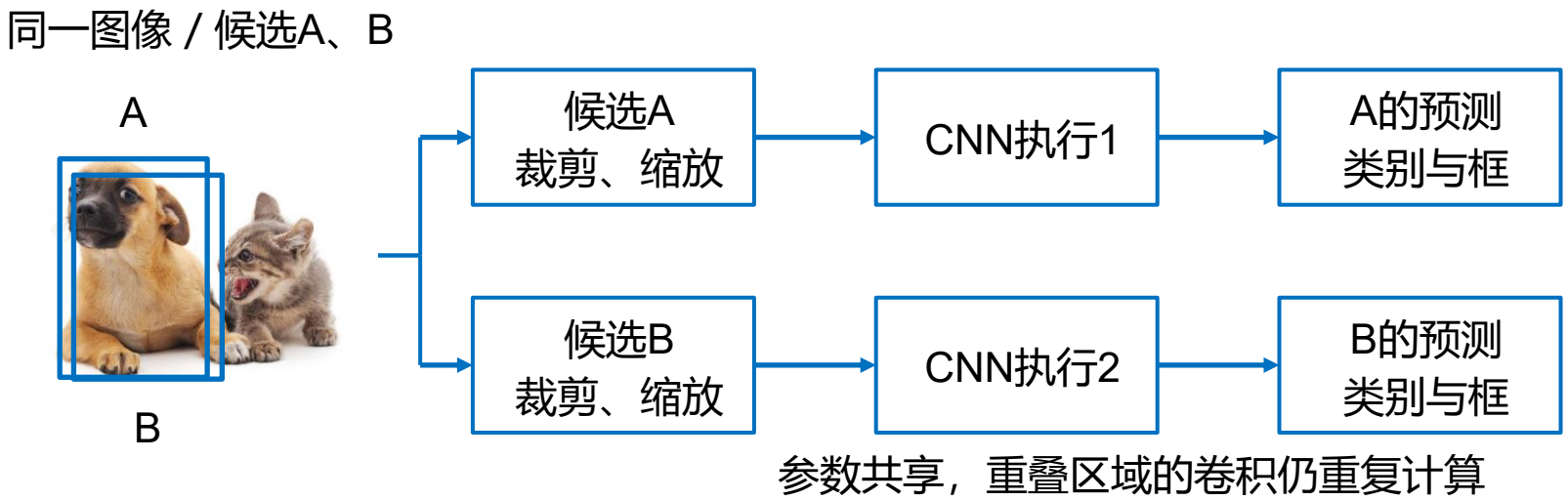
Selective Search　　每个候选分别执行　　可输出多个目标，也可没有目标

候选覆盖限制召回：漏掉的目标难以由后续识别补回；保留候选仍需分类、框回归与NMS

Girshick et al. Rich Feature Hierarchies. CVPR 2014.

R-CNN：对候选区域提取CNN特征

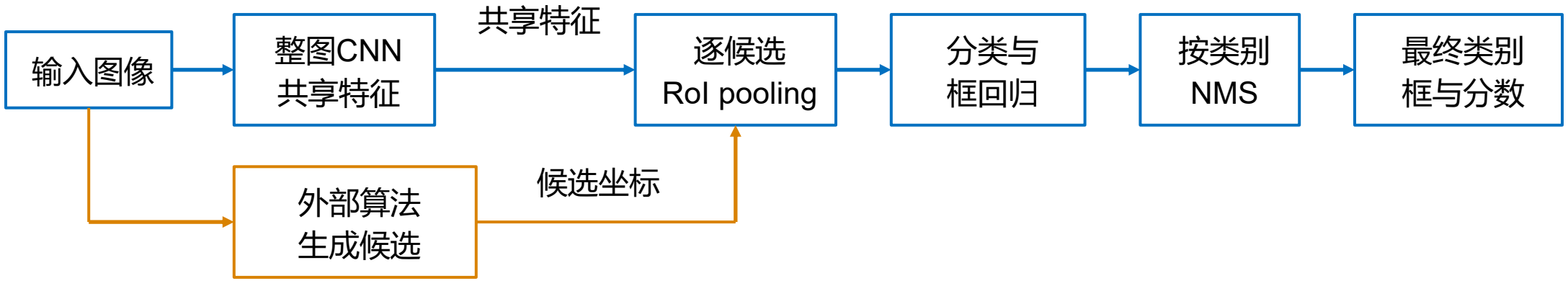
- 本图放大逐候选推理：裁剪、缩放、CNN特征、分类与框修正。
- A、B共用参数，仍各自卷积；预测经汇总与NMS后形成最终检测。
- 原始训练先分类预训练，再用检测数据微调CNN。
- 固定CNN特征后，分别训练SVM分类器与框回归器。



Girshick et al. Rich Feature Hierarchies. CVPR 2014.

Fast R-CNN：先共享特征，再按区域读取

- Fast R-CNN仍完成整图检测，采用一次共享CNN与按区域读取，减少逐候选卷积。
- 外部候选坐标和共享特征共同进入RoI pooling，再分类、修正框，并经NMS输出检测。
- 训练按RoI分配类别与框目标：分类损失用于正负样本，回归损失只用于正样本，联合优化网络。



RoI pooling：蓝线提供区域特征，橙线提供读取位置

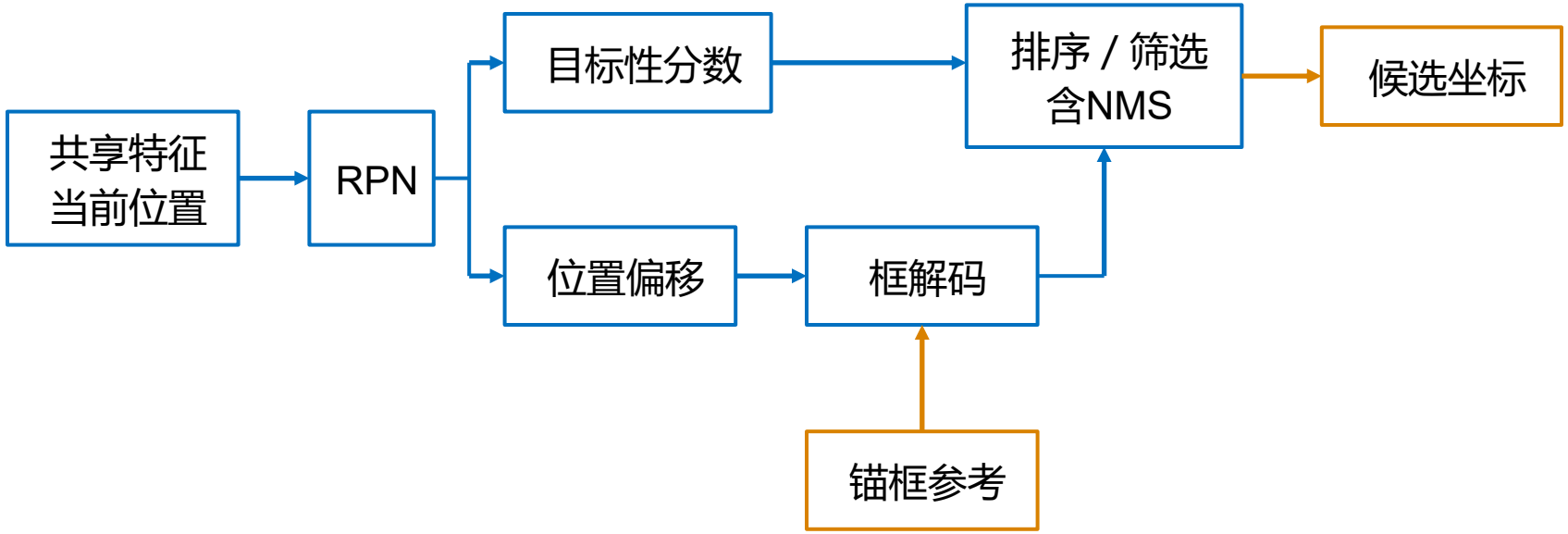
Girshick. Fast R-CNN. ICCV 2015.

3 用网络学习候选区域

整图特征已共享，候选仍由外部算法产生；用RPN学习候选

候选区域也可以由网络学习

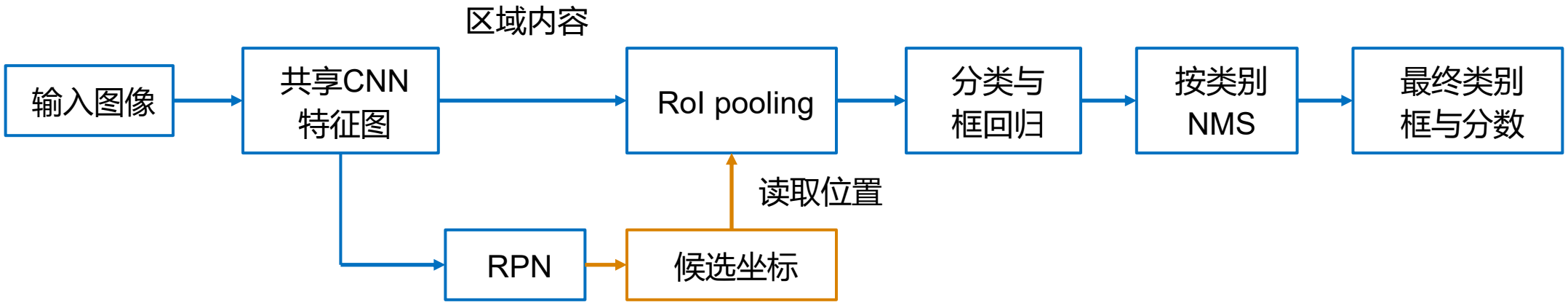
- Fast R-CNN已共享CNN计算，但候选仍由外部算法生成；RPN复用共享特征学习候选，将候选生成纳入网络。
- 对每个位置的锚框（预设参考框），预测目标性与位置偏移。
- 框解码后经排序与NMS产生候选坐标；具体类别由第二阶段判断。



Ren et al. Faster R-CNN. NIPS 2015.

Faster R-CNN的两阶段数据流

- 区域路线的变化：R-CNN逐候选算特征，Fast整图共享，Faster再学习候选。
- RoI pooling的两路输入：共享特征提供区域内容，RPN候选坐标指定读取位置。
- 训练时，RPN学目标 / 背景与候选框修正；检测头学具体类别 / 背景与框修正。
- 两阶段分别分配监督、计算分类与正样本回归损失，共同训练共享CNN。



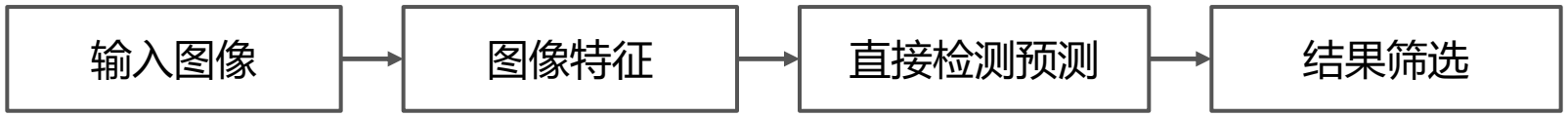
同一特征图同时用于候选生成与区域读取

4 单阶段与多尺度预测

Faster R-CNN仍按候选逐区域检测；比较直接预测与多尺度组织

单阶段：直接产生类别与位置预测

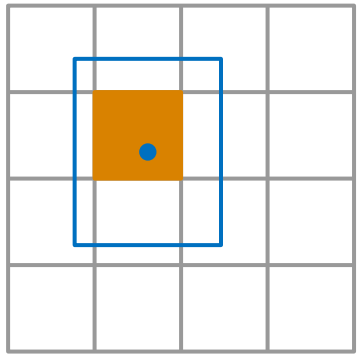
- 两阶段按候选区域再检测；单阶段直接预测类别与框，省去候选后的逐区域检测。
- 原始YOLO与SSD并列；多尺度、锚框与监督对应可组合，并非替代链。
- 稠密预测仍需分配位置监督、处理重复结果；速度与精度须按具体模型比较。



原始YOLO：从整图统一预测目标

- 原始YOLO用整图CNN统一预测类别与框，一次前向得到待筛选的预测。
- 中心所在网格负责目标；每格预测B个框及置信度，共享条件类别概率，同格多目标受限。
- 类别检测分数 = 条件类别概率×框置信度，再筛选和NMS；置信度兼顾目标存在与定位质量。



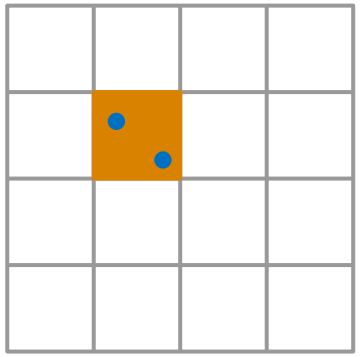


中心所在单元负责目标

每格：B个框及置信度

每格共享一组类别概率

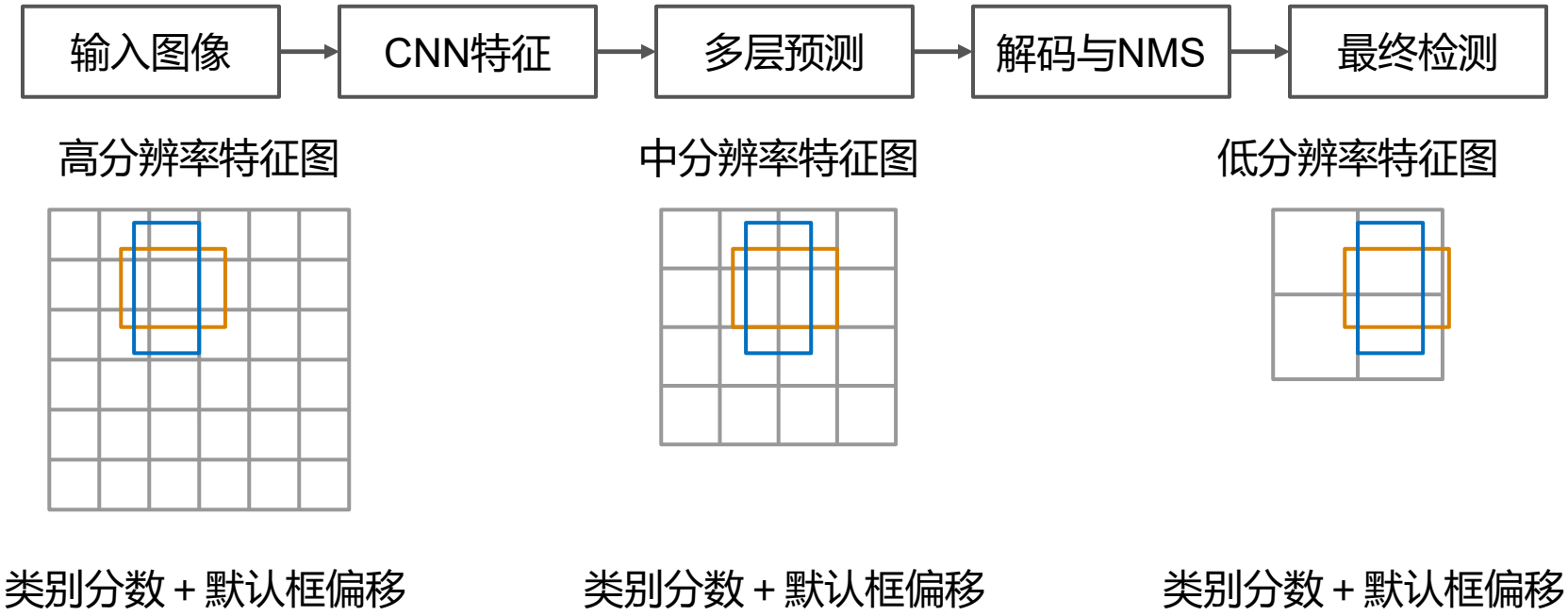
两个中心落入同一格



责任冲突限制密集目标

SSD：在多个特征图上直接预测

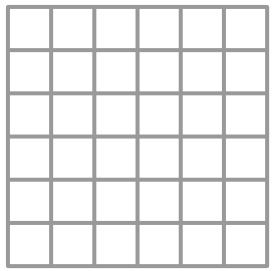
- SSD用多层特征同时寻找不同大小的目标，在各层的空间位置直接预测类别与框。
- 每位置配置不同形状的**默认框**并预测类别和偏移；训练用IoU分配监督，推理汇总各层、解码并执行NMS。



多尺度预测怎样帮助覆盖不同目标

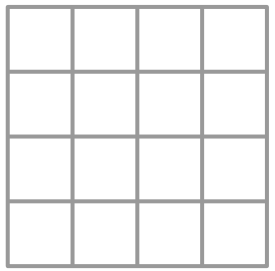
- 高分辨率保留较密的空间细节；较深层通常感受野更大、语义更抽象，多层提供不同的定位与识别信息。
- SSD配合预测层与默认框尺度；多尺度与锚框是可组合的维度，小目标仍需分辨率和语义质量。

高分辨率



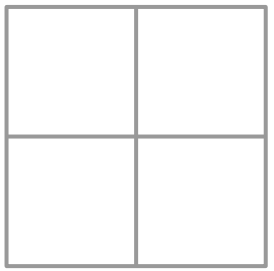
密集位置，保留细节
配较小默认框
有助于小目标定位

中分辨率



空间细节与上下文
配中等默认框
补充中尺度目标

低分辨率



通常感受野更大
深层语义更抽象
配较大默认框

SSD：多层分别预测；FPN：跨层融合，为高分辨率补充深层语义

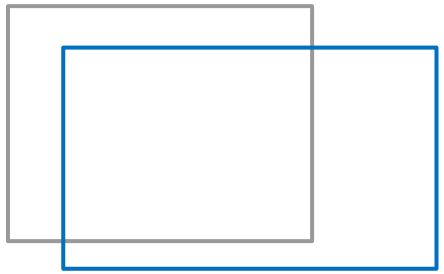
5 框表示与目标集合预测

SSD仍以默认框为参考，也需筛除重复；比较位置表示与集合一对一对应

为什么考虑去掉预设锚框

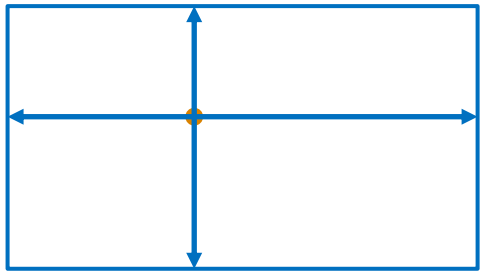
- 锚框以预设位置和形状为参考；尺度、长宽比、数量和匹配规则都需要设计。
- **无锚框检测**可从位置直接表示边界，减少预设形状依赖；是否有锚框与是否两阶段是不同维度。

锚框：参考形状 + 偏移



灰：参考框；蓝：预测框

FCOS：位置 + 四边距离



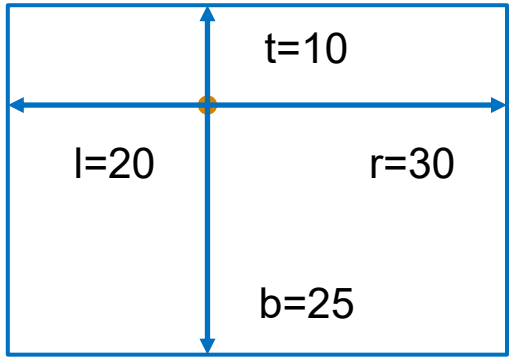
从内部位置恢复边界

FCOS：从特征位置预测四个边界距离

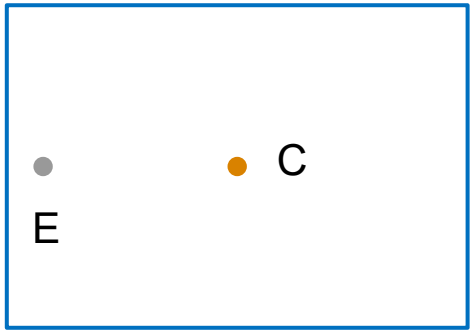
- FCOS把框表示为某个位置到目标四边的距离，直接从位置预测，省去预设锚框。
- 每位置输出类别与四边距离，再恢复框；中心度依四边距离构造监督并调节分数，它不是IoU。仍需监督分配和NMS。



位置 (50,40)



框：(30,30) 到 (80,65)



示意位置中心度：C=1，E≈0.30

C=(55,47.5)，E≈(33.94,47.5)

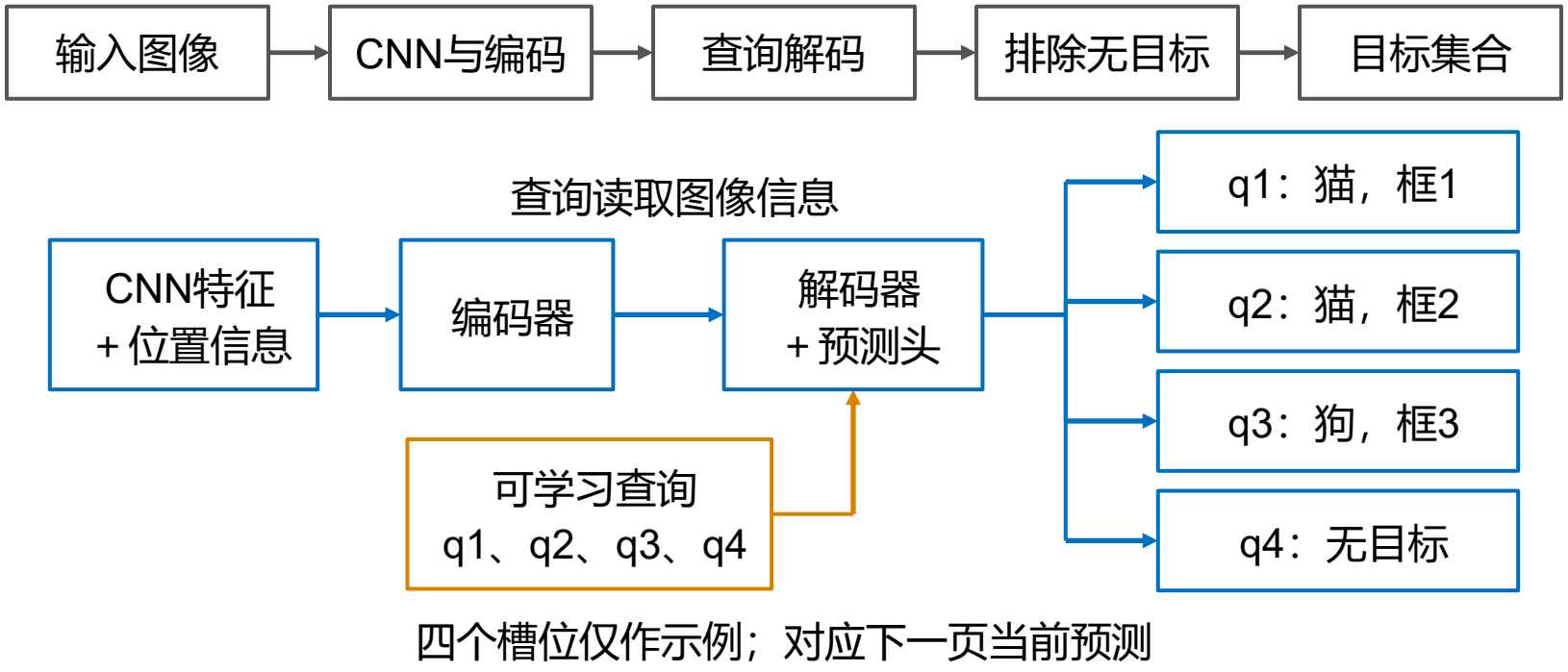
类别分数0.8；乘图示中心度得0.80 / 0.24

直接预测目标集合的思路

- FCOS改变框表示，仍有重复预测；**DETR**用查询槽位与一对一监督，直接预测目标集合。
- 每个槽位输出类别与框（或无目标）；槽位数固定，目标数量由无目标判断改变。
- 训练用一对一匹配，让每个真实目标只监督一个槽位，其余学习“无目标”，改变稠密位置的多预测监督关系。
- 解码器让查询读取图像并交换信息；原始DETR收敛慢，小目标检测较弱。

DETR：用查询组织一组目标预测

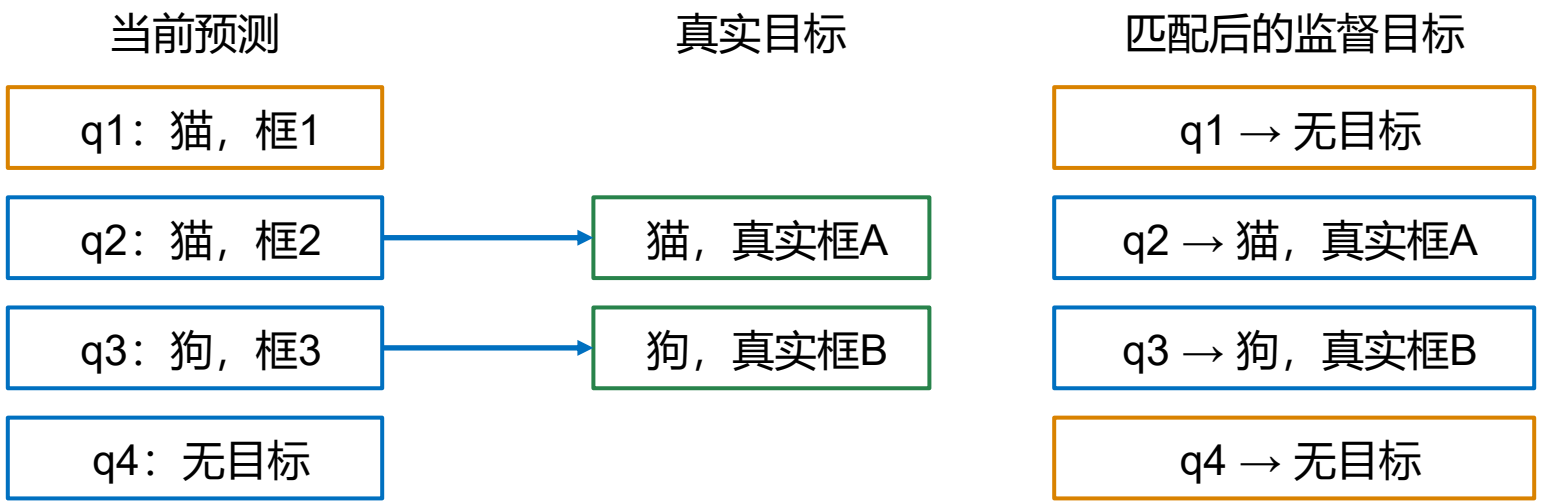
- DETR先把图像编码成可读取的特征，再让一组查询从中提取目标信息。
- 带位置信息的CNN特征经编码后送入解码器；各查询输出框及类别或无目标，编号不固定对应目标。



Carion et al. End-to-End Object Detection with Transformers. ECCV 2020.

一对一匹配怎样改变重复处理

- 训练先按当前类别和位置代价求一对一匹配，确定监督对象；蓝线为选中关系。
- 匹配后，q2 / q3计算类别与框损失；q1 / q4只计算“无目标”类别损失，再反向传播优化参数。
- DETR推理排除“无目标”，不运行匹配或传统NMS；仍可能重复或误检。
- 两个变化分别是：FCOS改变框表示，DETR改变监督对应。



匹配选对应，损失更新参数；q1学无目标以降低重复猫预测的得分

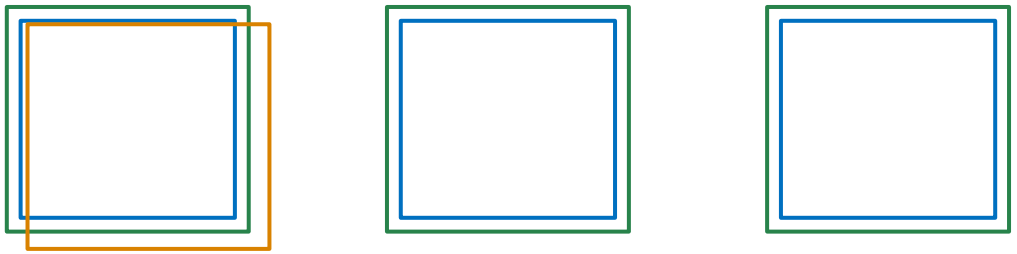
6 检测评价与方法比较

框与集合的预测都可能出错；用检测评价和共同设计问题比较结果与方法

怎样判断一次检测是否正确

- 训练匹配分配学习目标；评价匹配统计检测正确性，两者目的不同。
- 本例按分数匹配同类目标：IoU达到0.5且目标未匹配时计真阳性（TP）；错误或重复计假阳性（FP），漏检计假阴性（FN）。
- $\text{Precision} = \frac{TP}{TP+FP}$，$\text{Recall} = \frac{TP}{TP+FN}$。绿色为标注真实框。

同一类的3个真实目标；绿色为真实框



0.95：TP

0.90：重复FP

0.80：TP

0.60：TP

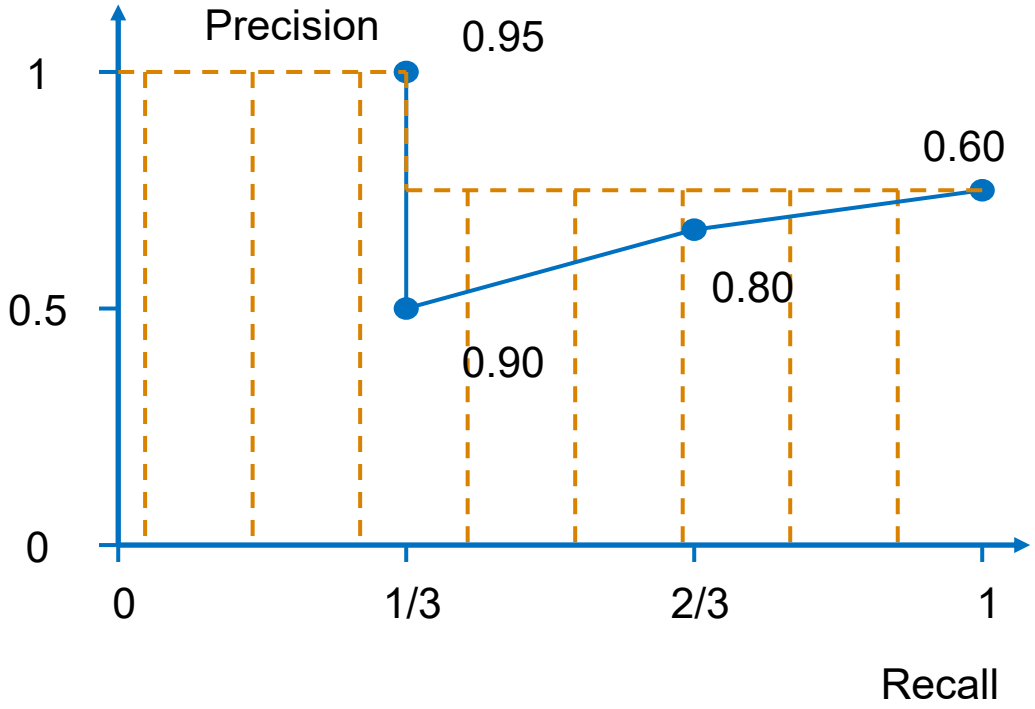
阈值0.85：TP=1，FP=1，FN=2

Precision = 1/2

Recall = 1/3

AP与方法比较需要哪些条件

- 每个召回率处取该点及右侧最高精度，形成插值包络；AP沿召回率方向汇总其下方面积，按协议离散计算。
- AP50固定IoU阈值0.50；COCO主AP对0.50至0.95（步长0.05）及类别平均。比较须统一数据、输入和计时条件。



降低阈值，预测依次加入

分数	Precision	Recall
0.95	1	1/3
0.90	1/2	1/3
0.80	2/3	2/3
0.60	3/4	1

蓝：PR点 / 加入顺序；橙：右侧最高精度
橙色竖线区域：沿Recall方向汇总的面积

从设计问题回看方法演进

方法	预测从哪里产生	框如何表示	训练怎样对应目标
R-CNN / Fast	外部候选区域	候选框 + 回归修正	按区域IoU分配监督
Faster R-CNN	RPN候选 + 区域检测	锚框 / 候选框 + 偏移	两阶段分别按IoU分配
原始YOLO	网格单元	网格相对中心、宽高	中心单元及负责框
SSD	多层位置的默认框	默认框 + 偏移	IoU匹配，可多框对应一目标
FCOS	多层稠密位置	位置 + 四边距离	框内位置及尺度等规则
原始DETR	固定查询槽位	直接输出中心、宽高	集合一对一，其余无目标

共享计算的变化：R-CNN逐候选执行CNN，Fast R-CNN先计算整图特征，再按候选区域读取。

三个问题贯穿各路线；特征共享、多尺度、有无锚框与阶段数是可组合的设计选择。

本讲总结

- 检测以实例类别与边界框组成输出，需要处理数量、位置、尺度与背景变化。
- 区域检测从候选区域识别目标，共享特征与学习候选分别改善不同处理环节。
- 单阶段检测直接产生预测，多尺度组织支持不同大小的目标。
- 无锚框方法改变位置表示，集合预测进一步改变预测与目标的对应关系。
- 方法演进围绕具体困难与取舍展开，检测质量和速度须在明确条件下比较。