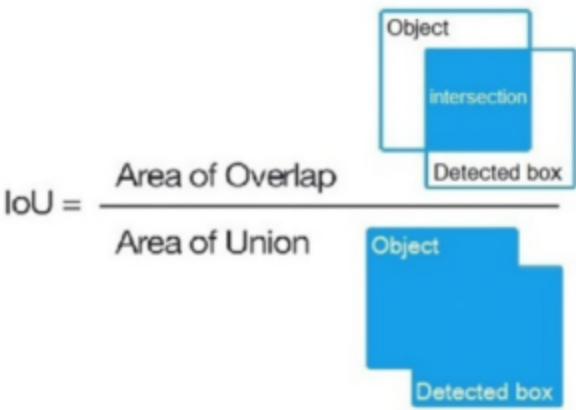


平均精度 mean Average Precision (mAP) (AP)

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

交并比 Intersection over Union (IoU)



真阳性 True Positive (TP)

假阳性 False Positive (FP)

假阴性 False Negatives (FN)

只要检测出来了，不管IoU，都是positive，大于IoU的才是True，小于IoU阈值的是False

没检测出来，就是Negatives，第二个框都没有，IoU就是0,所以是False

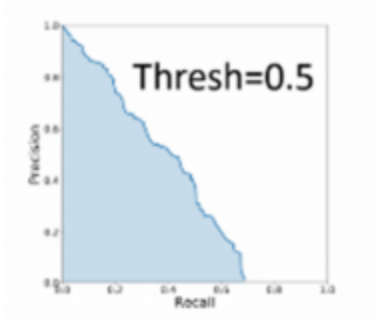
召回：真实存在，检测出来多少

召回  $Recall = \frac{TP}{TP + FN} = \frac{TP}{\# \text{ ground truths}}$

精度：所有检测出来的，真实的多少

精度  $Precision = \frac{TP}{TP + FP} = \frac{TP}{\# \text{ predictions}}$

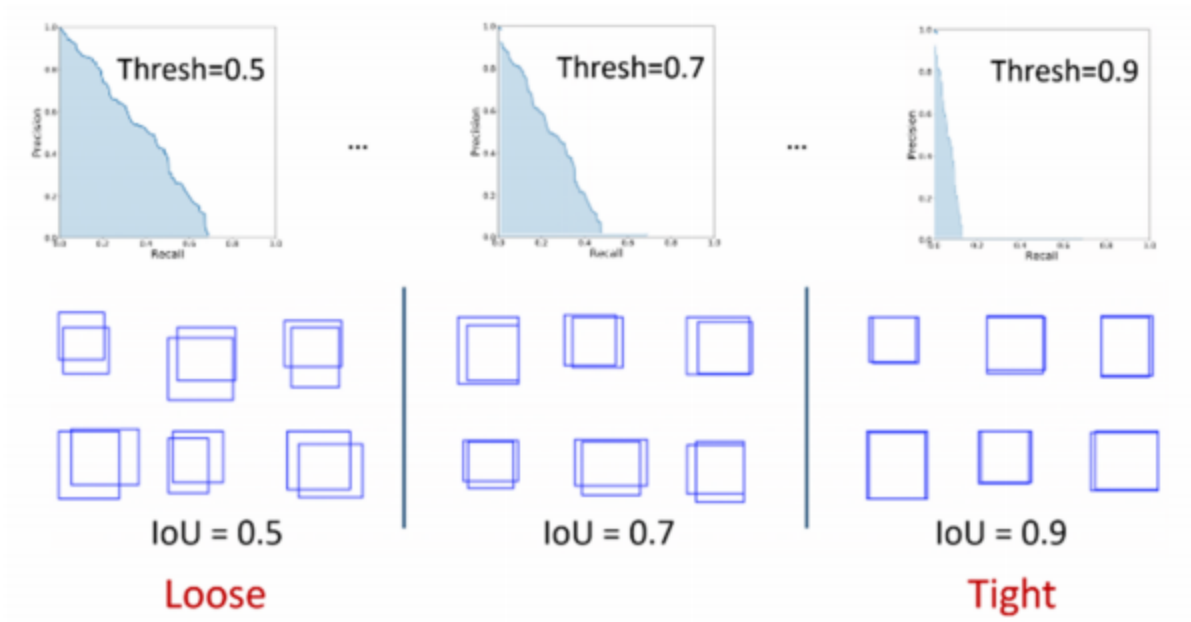
AP50 or AP0.5



精度召回曲线下的面积

AP0.50: 0.55: 0.95

$$mAP_{COCO} = \frac{mAP_{0.50} + mAP_{0.55} + \dots + mAP_{0.95}}{10}$$





MOTA (Multiple Object Tracking Accuracy, 多目标跟踪准确度) 是评估多目标跟踪系统性能的一个指标。公式如下所示:

$$MOTA = 1 - \frac{FN + FP + IDSW}{GT}$$

以下是公式中每一个符号的详细解释:

#### 1. FN (False Negatives):

- 假阴性, 指跟踪系统未能检测到的目标数。即存在真实的目标 (ground truth), 但跟踪系统没有捕捉到它们。
- 举例: 假设场景中有5个真实目标 (GT=5), 但系统只检测到4个目标, 漏掉了1个真实目标, 则  $FN = 1$ 。

#### 2. FP (False Positives):

- 假阳性, 指跟踪系统错误地检测到的目标数。即跟踪系统检测到了某些目标, 但这些目标实际上并不存在。
- 举例: 场景中只有5个真实目标, 但系统检测到7个目标, 其中2个是错误的, 则  $FP = 2$ 。

#### 3. IDSW (Identity Switches):

- 身份切换错误, 指跟踪系统在连续帧中错误地分配了目标的身份标签。例如, 目标 A 和目标 B 的身份标签在跟踪过程中互换。
- 举例: 系统在第一帧中将目标 A 赋值为 ID1, 目标 B 赋值为 ID2, 但在第二帧中误将 A 标记为 ID2, B 标记为 ID1, 则  $IDSW = 1$ 。

#### 4. GT (Ground Truth):

- 真实目标的数量, 表示场景中所有真实存在的目标数。
- 举例: 如果场景中共有5个真实目标, 那么  $GT = 5$ 。

#### 举例说明:

假设某场景中有  $GT = 5$  个真实目标。

- 系统漏掉了1个目标 ( $FN = 1$ )。
- 系统多检测了2个不存在的目标 ( $FP = 2$ )。
- 在跟踪过程中, 系统发生了1次身份切换错误 ( $IDSW = 1$ )。

则按照公式计算:

$$MOTA = 1 - \frac{FN + FP + IDSW}{GT} = 1 - \frac{1 + 2 + 1}{5} = 1 - \frac{4}{5} = 0.2$$

因此, 该场景的 MOTA 为 0.2, 表示跟踪系统的准确率为 20%。



# Performance Metrics

性能指标

- PCK (Percentage of Correct Keypoint) 正确关键点百分比
  - Measure the percentage of predicted keypoint that falls within a normalized distance of true joint, e.g., 测量预测关键点落在真关节归一化距离内的百分比，例如：
  - In Flic dataset, PCK@0.2: Distance between predicted and true joint  $< 0.2 \times$  torso diameter. 在Flic数据集中，PCK@0.2: 预测关节与真实关节之间的距离 $< 0.2 \times$ 躯干直径。
  - In MPII dataset, PCKh@0.5 refers to the threshold = 50% of the head segment line. 在MPII数据集中，PCKh@0.5表示阈值=头部线段线的50%。
- PCP (Percentage of Correct Parts) PCP (正确零件百分比)
  - Measure the localization accuracy of limbs. 测量肢体定位精度。
  - If the distance between the two predicted joints and the true limb is less than a fraction of the limb length (e.g., 0.1-0.5), then the limb is considered detected correctly.

如果两个预测关节与真实肢体之间的距离小于肢体长度的一小部分（例如0.1-0.5），则认为该肢体已被正确检测。

## Performance Metrics

性能指标

- AP and AR based on OKS (Object Keypoint Similarity) 基于OKS (对象关键点相似度) 的AP和AR
  - Average Precision (AP), Average Recall (AR) and their variants are used in COCO, MPII and PoseTrack. 平均精度 (AP)、平均召回率 (AR) 及其变体在COCO、MPII和oserack中得到了应用。
  - Object Keypoint Similarity (OKS) plays the same role as the IoU in detection. 对象关键点相似度 (Object Keypoint Similarity, OKS) 在检测中的作用与IoU相同。
  - OKS definition: the distance between predicted points and ground truth joints normalized by the scale of the person. OKS定义: 由人的尺度归一化的预测点与地面真值节点之间的距离。

Metric of COCO dataset with OKS: 带有OKS的COCO数据集度量:

```
Average Precision (AP):
AP          % AP at OKS=.50:.05:.95 (primary challenge metric)
APOKS=.50   % AP at OKS=.50 (loose metric)
APOKS=.75   % AP at OKS=.75 (strict metric)
Average Recall (AR):
AR          % AR at OKS=.50:.05:.95
AROKS=.50   % AR at OKS=.50
AROKS=.75   % AR at OKS=.75
```



## 1. PCK (Percentage of Correct Keypoints)

- **定义：** PCK表示预测关键点与真实关键点之间的归一化距离小于某个阈值时的百分比。
- **关键点距离归一化：**
  - 在 FLIC 数据集中，使用躯干直径的 0.2 倍作为阈值（PCK@0.2）。
  - 在 MPII 数据集中，使用头部线段长度的 50% 作为阈值（PCKh@0.5）。
- **公式：**

$$PCK = \frac{\text{预测关键点与真实关键点距离小于阈值的点数}}{\text{总关键点数}}$$

**例子：**

- 场景中有10个真实关键点。
- 系统预测出8个关键点，其距离真实关键点分别为：0.1、0.2、0.15、0.3、0.1、0.25、0.2、0.1。
- 阈值为0.2（归一化距离）。
- 预测点中有6个距离小于0.2的真实点， $PCK = 6/10 = 60\%$ 。

## 2. PCP (Percentage of Correct Parts)

- **定义：** PCP表示预测肢体的定位精度。若预测的肢体端点与真实端点的距离均小于肢体长度的一部分（如0.1到0.5），则该肢体被认为是正确检测的。
- **公式：**

$$PCP = \frac{\text{正确检测到的肢体数}}{\text{总肢体数}}$$

**例子：**

- 假设一幅图像中有5条真实肢体。
- 系统检测到4条肢体，其中3条的端点距离小于肢体长度的0.2，1条距离大于阈值。
- 正确检测的肢体数为3， $PCP = 3/5 = 60\%$ 。

## 3. 基于OKS的AP和AR

- **OKS (Object Keypoint Similarity)：**
  - 对象关键点相似度，类似于目标检测中的IoU（交并比）。
  - **定义：** 预测点与真实点之间的距离，归一化为人体的尺度（例如头部宽度或躯干长度）。
  - **公式：**
$$OKS = \exp\left(-\frac{d^2}{2s^2k^2}\right)$$
    - $d$ ：预测点与真实点之间的欧几里得距离。
    - $s$ ：人体的比例尺度。
    - $k$ ：关键点的可见性权重。
- **AP (Average Precision)：**
  - 基于OKS的平均精度，表示在不同OKS阈值下预测的关键点的准确性。
  - 示例： $AP@OKS = 0.50 : 0.05 : 0.95$  是在从0.5到0.95范围内均匀取样多个阈值计算的平均值。
- **AR (Average Recall)：**
  - 基于OKS的平均召回率，表示在不同OKS阈值下预测的关键点的召回情况。

**例子：**

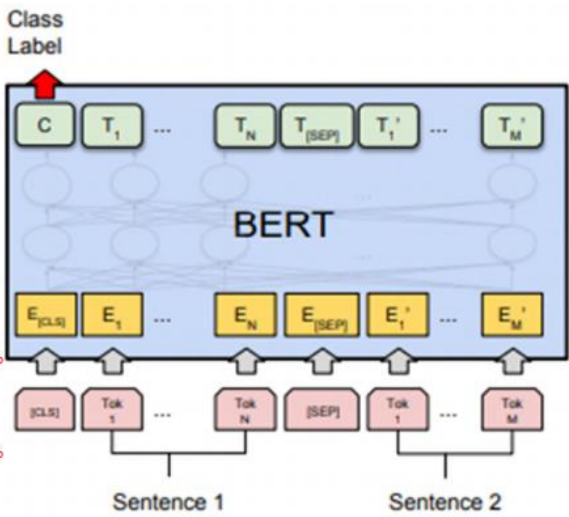
- 一个COCO数据集中，有100个真实关键点。
- 系统预测出80个关键点，其中60个点的OKS > 0.5。
- $AP@0.5 = 60/100 = 60\%$ 。



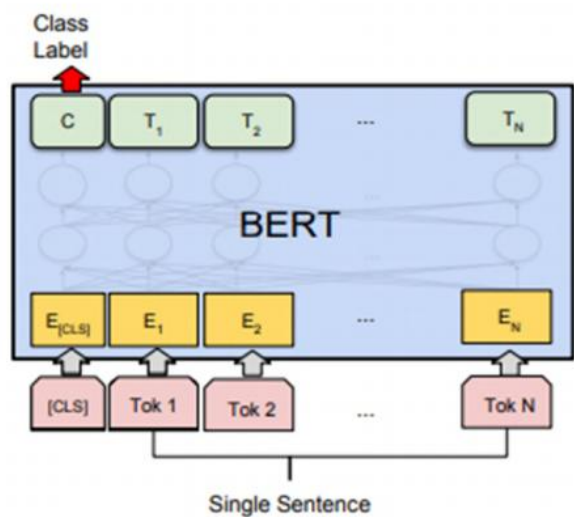


# Performance Evaluation 性能评估

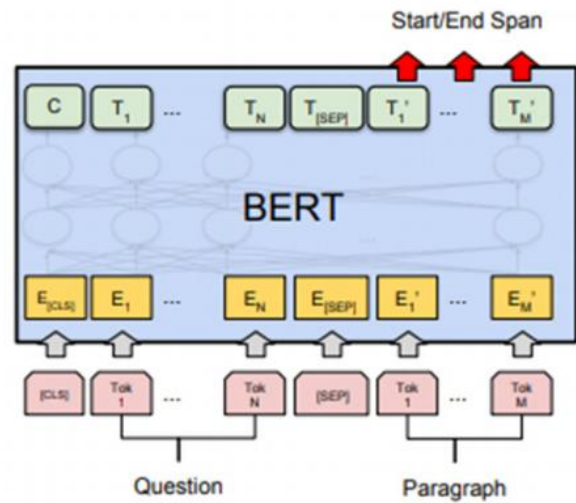
- General Language Understanding Evaluation (GLUE) 一般语言理解评估 (GLUE)
  - Sentence pair task. 句子配对任务。
  - Single sentence classification task. 单句分类任务。
- Stanford Question Answering Dataset (SQuAD) 斯坦福问答数据集 (SQuAD)



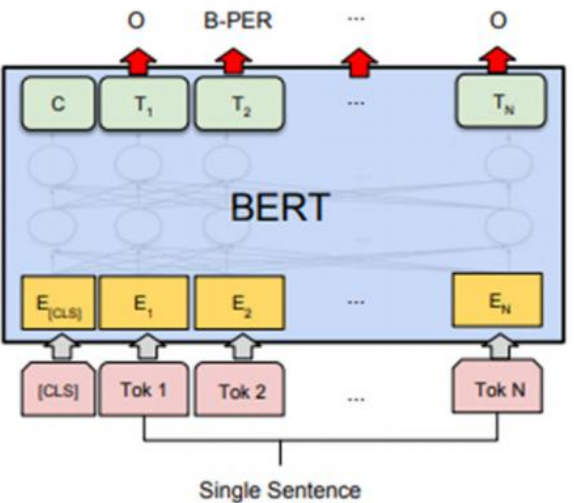
(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG



(b) Single Sentence Classification Tasks: SST-2, CoLA



(c) Question Answering Tasks: SQuAD v1.1



(d) Single Sentence Tagging Tasks: CoNLL-2003 NER