## 基本原理：

### 导言

该论文介绍了一种新颖的物体检测方法，该方法将目标检测看作是一个直接的集合预测问题。这种方法称为Detection TRansformer(DETR)，它消除了许多手工设计的组件的需求，例如非最大抑制(NMS)和锚点生成，而这些组件通常用于传统的物体检测方法。

### DETR 如何进行目标检测

**直接集合预测**：DETR没有使用涉及区域建议网络（RPN）和后续对象分类的传统两阶段过程，而是将对象检测构建为直接集合预测问题。它将图像中的所有对象视为一个集合，旨在一次性预测它们的类和边界框。

**对象查询**：DETR引入了“对象查询”的概念。这些查询表示模型需要预测的对象。无论图像中的对象数量如何，对象查询的数量通常是固定的。

**Transformer Self-Attention：**该机制应用于对象查询和从输入图像中提取的空间特征（称为键和值）。这种自注意力机制使DETR能够学习物体与其空间位置之间的复杂关系和依赖关系。

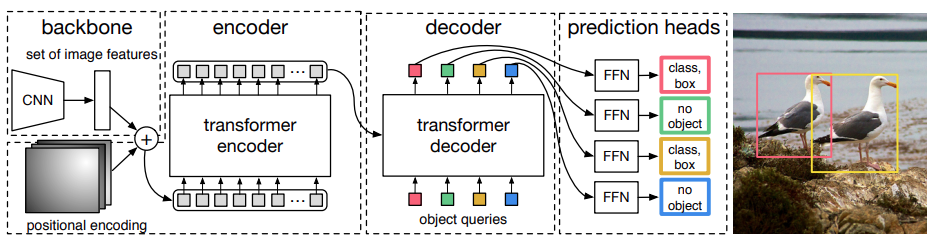
**并行预测**：使用从自注意力机制收集的信息，DETR同时预测每个对象查询的类和位置（边界框）。这种并行预测与传统的物体检测器不同，因为传统的物体检测器通常依赖于顺序处理。

**二分匹配**：为确保每个预测的边界框与图像中的真实对象相对应，DETR使用二分匹配将预测结果与实际结果相关联。此步骤可提高模型在训练期间的精度。

### DETR 如何工作？

DETR中有两个部分对于检测中的直接集合预测是必不可少的：(1) 迫使预测和真实框之间进行唯一匹配的集合预测损失；(2)预测一组对象并建模它们之间关系的体系结构。

1. DETR的集合预测损失通过Hungarian算法实现了预测集合与真实集合之间的一一对应匹配，并结合分类损失和边界框回归损失，构建了一个端到端的物体检测模型。这个方法不仅简化了检测流程，还提升了检测精度。
2. 整体DETR架构如下所示。



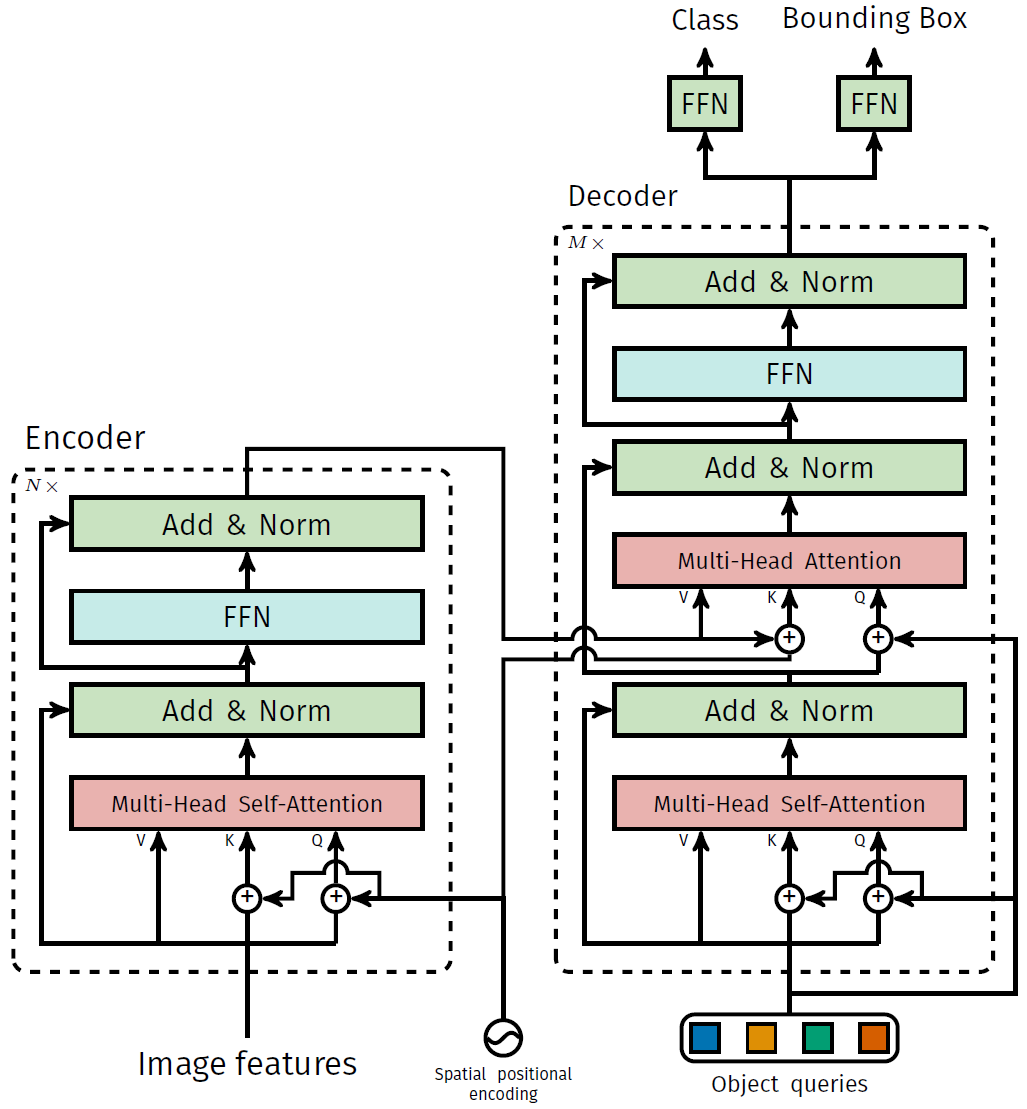
DETR使用传统的CNN骨干来学习输入图像的二维表示。该模型对其进行扁平化，并辅以位置编码，然后将其传入transformer编码器。然后transformer解码器将少量固定数量的学习到的位置嵌入(我们称之为对象查询)作为输入，并额外关注编码器输出。我们将解码器的每个输出嵌入传递给一个共享的前馈网络(FFN)，该网络可以预测检测结果（类别和边界框)或一个‘无对象’类别。

**Backbone:**

DETR首先通过卷积神经网络（CNN）处理输入图像,从图像中提取高级特征表示。这些特征保留了有关图像中对象的空间信息，并作为后续操作的基础。

**Transformer:**

DETR中transformer的架构如下：

****

在DETR中，Transformer 的编码器用于处理输入图像的特征，并将其转换成编码表示。在这个过程中，引入了固定的位置编码，它们被添加到CNN的输出中。为模型提供了有关图像不同部分之间的空间关系的信息。

解码器使用多头自注意力机制将大小为d的N个嵌入进行转换。在每个解码器层中，并行地解码N个对象，而不是使用自回归模型逐个预测输出序列的元素。

同时，通过全局地考虑所有目标对象之间的关系，模型能够更好地理解整个图像的上下文信息，从而提高检测准确性。

**预测前馈网络（FFNs）：**

最终的预测由一个具有ReLU激活函数和隐藏维度d的3层感知器和一个线性投影层来完成。FFN预测输入图像的归一化中心坐标、高度和宽度，线性层使用softmax函数预测类标签。

由于模型预测的是一个固定大小的边界框集合，而这个集合大小N通常远大于图像中感兴趣对象的实际数量，因此使用了一个额外的特殊类别标签来表示在某个位置未检测到对象。

### DETR的缺点

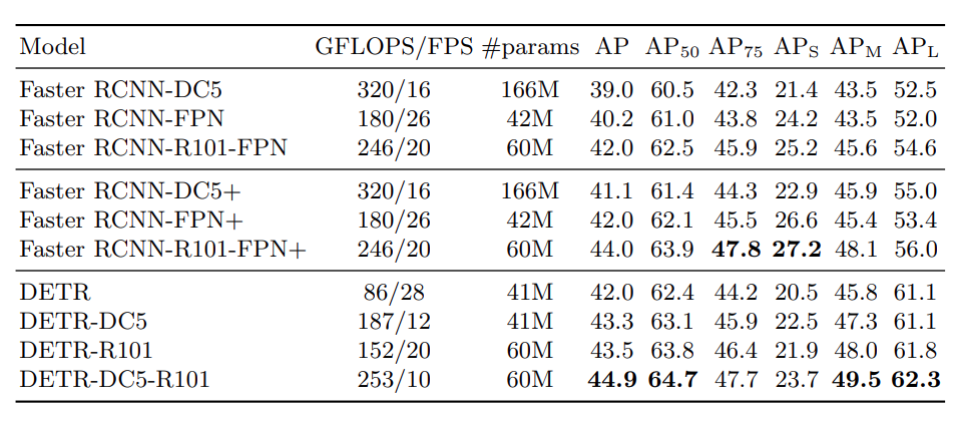
**高计算资源**：训练和使用DETR可能需要大量计算，尤其是对于大型模型和高分辨率图像。这可能会限制无法访问强大硬件的研究人员和从业人员的可访问性。

**固定对象查询计数**：DETR需要预先指定对象查询的数量，这在处理包含不同数量对象的场景时可能是一个限制。不正确的查询数可能会导致漏检或效率低下。

**在小物体上性能较低**：DETR在大型对象上显示了显著更好的性能，这一结果很可能是由transformer的非本地计算所实现的。但它在小物体上的性能较低。

### DETR的性能

在最初发布时，作者的主要目标是通过DETR超越Faster R-CNN基线。然而，在目前的情况下，与DETR型号相比，增强的Faster R-CNN ResNet50 FPN V2 表现出卓越的性能。



## 应用场景：

医学影像分析：DETR可以用于医学影像分析，帮助医生识别和定位图像中的病变区域，提高医学诊断的准确性和效率。

物体检测与识别：DETR可以用于从图像中检测和识别各种物体，例如汽车、人、动物等。这在自动驾驶、监控系统、智能家居等领域都有广泛的应用。

视频分析：DETR也可以应用于视频分析领域，帮助实现视频中物体的检测、跟踪和识别，从而用于视频监控、行为分析等应用。

## 最新成果：

RT-DETR: DETRs Beat YOLOs on Real-time Object Detection

如图所示

