

## RTMPose 简要介绍

**RTMPose**，全称为 Real-Time Multi-Person Pose Estimation based on MMPose。即基于 MMPose 的实时多人姿态估计算法。它的优点有：出色的实时性、较高的准确性、灵活的部署性、易于开发和使用以及对复杂场景的适应性。

**工作原理如下：**

总体架构：该算法采用自上而下（top-down）的方法，先使用现有的检测器获得人体检测边界框，然后单独估计每个人的姿态。该方法虽然准确但速度较慢，得益于高效的实时检测器，人体检测部分不再是 top-down 方法推理速度的瓶颈。

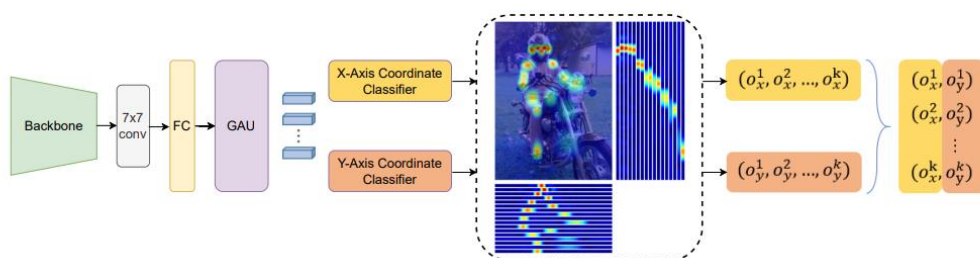


图 1. 模型架构

(2) 骨干网络：选择 CSPNeXt 作为骨干网络，CSPNeXt 在速度和准确性之间取得了比较好的平衡，且易于部署。

(3) 关键点定位方法：使用基于 SimCC (Simultaneous Confidence Maps for Classification and Coordinate Regression) 的算法预测关键点。该算法将关键点定位公式化为一个分类问题，其核心思想是将水平轴和垂直轴划分为 equal-width numbered bins，并将连续坐标离散为积分 bin 标签。然后对模型进行训练，以预测关键点所在的 bin，可以通过使用大量的 bin 来将量化误差减少到 subpixel level。与基于热图 (heatmap) 的算法相比，以更低的计算量实现了具有竞争力的精度。

**性能方面：**RTMPose-m 在 COCO 数据集上实现了 75.8% 的 AP，并且在 Intel i7-11700 CPU 上达到 90+ FPS，在 NVIDIA GTX 1660 Ti GPU 上达到 430+ FPS。RTMPose-s 模型在 Snapdragon 865 芯片上以 70+ 的 FPS 在 COCO 上实现了 72.2% 的 AP，优于现有的开源库。

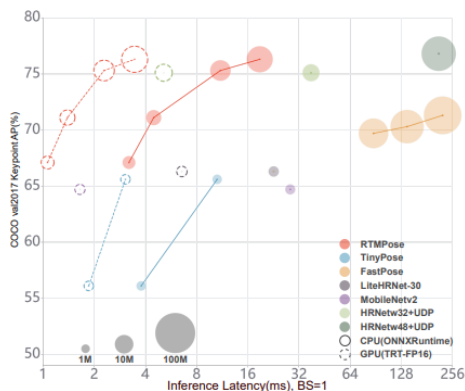


图 2. RTMPose 和其他开源模型在 COCO val set 上关于模型大小、延迟和精度的比较。圆形大小表示模型参数的相对大小。