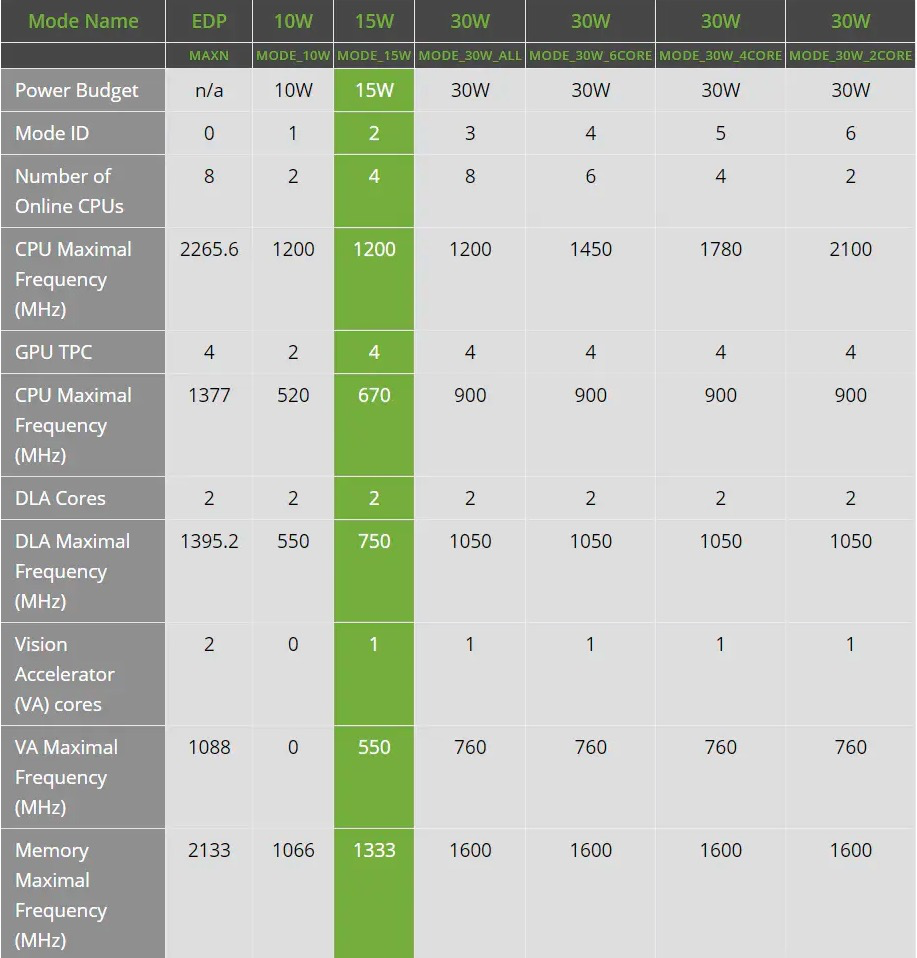
# BM-BEV在Xavier上的性能评估

# Xavier参数

|  |  |
| --- | --- |
| 算力 | 32 TOPs |
| GPU | 512 核 Volta GPU (具有 64 个 Tensor 核心) 11 TFLOPS (FP16) 22 TOPS (INT8) |
| CPU | 8 核 ARM v8.2 64 位 CPU、8 MB L2 + 4MB L3 |
| 功耗 | 10W/15W/30@ |

更多参数请参考：[NVIDIA Jetson AGX Xavier 提供 32 TeraOps，为机器人技术展开一个新时代](https://developer.nvidia.com/zh-cn/blog/nvidia-jetson-agx-xavier-32-teraops-ai-robotics/)

# Xavier工作模式



默认是Mode 2，后续性能评估会让Xavier满状态运行，即**切换到Mode 3**

# Float模型性能评估

## FPS

计算方式为：

import time  
  
total\_time = 0.0  
for i, data in enumerate(data\_loader):  
 torch.cuda.synchronize()  
 start = time.time()  
 model(\*\*data)  
 torch.cuda.synchronize()  
 end = time.time()  
 total\_time += (end - start)  
  
batch\_size = len(data\_loader)  
print('FPS: ', 1 / (total\_time / batch\_size))

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Flops** | **Backbone** | **FPS** |
| BM-BEV (with Lidar) | / | EfficientNet-B3 | 0.55 |
| BM-BEV (wo Lidar) | / | EfficientNet-B3 | 1.4 |

BM-BEV-wo-Lidar的FPS大概是BM-BEV-with-Lidar的**3倍**。

## Profile

这里使用了Pytorch自带的性能分析工具[Profiler](https://pytorch.org/tutorials/beginner/profiler.html)来对BM-BEV中的OP的耗时进行分析。

### BM-BEV-with-Lidar



可以看到Vox计算在整个inference时间占比比较大，且无其他运算并行，导致整体inference效率较低。

### BM-BEV-wo-Lidar



这两个红框处有较多的非卷积操作，

其中，红框1:



红框2:



# 附：在Chrome中打开Profile可视化文件

首先，在浏览器一栏输入：chrome://tracing/

然后，点击Load选择生成的profile文件即可。

