Energy Consumption and Energy Efficiency

Accurately predicting energy consumption requires the right metrics. However, currently, metrics of this sort are not standardized, and the closest metric to experiment with is Floating Point Operations (FLOPs). It's important to note that this will be an approximation.

Here's a rough method to estimate power consumption:

1. **Determine the FLOP per watt (FLOP/W) of the hardware (GPU)**: This value varies significantly between different types of hardware
2. **Calculate the total number of FLOPs**: We already have this from our previous calculation.
3. **Estimate power consumption**: Divide the total number of FLOPs by the FLOP/W to get the power consumption in watts.

1. FLOPS/W

Efficiency is defined as the number of **FLOPS per Watt**.

Base on our research, datasheet, and calculate, Nvidia A100 GPU has the energy efficiency below:

Floating Point format	Nvidia A100 for Nvidia HGX (Max 400W)	Nvidia A100 for PCle (Max 250W)
FP8	N/A	N/A
FP16	0.78 TFLOPS/W	1.248 TFLOPS/W
FP32	0.04875 TFLOPS/W	0.078 TFLOPS/W
FP64	0.02425 TFLOPS/W	0.0388 TFLOPS/W

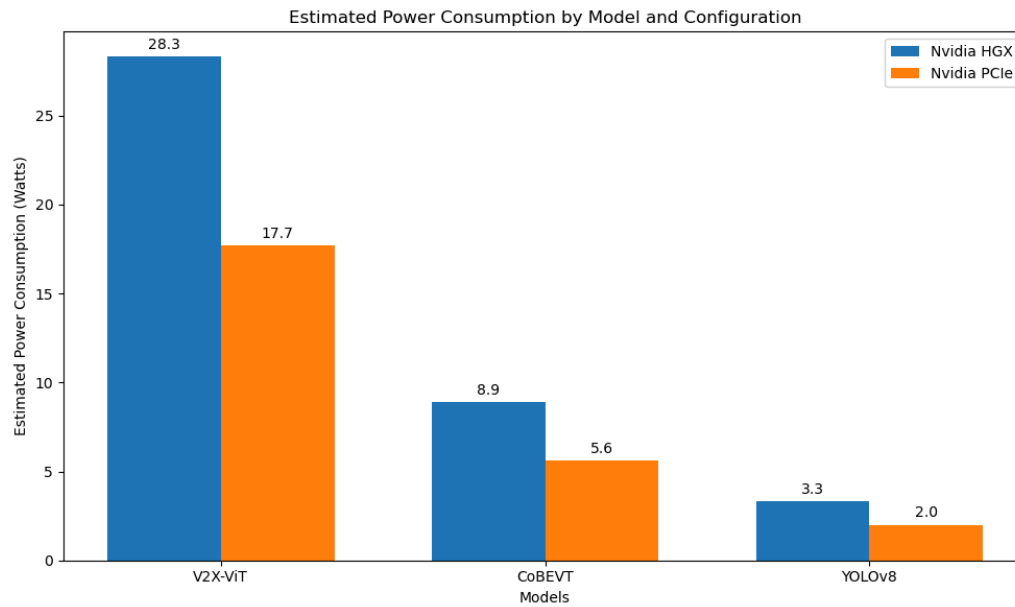## 2. Total number of FLOPs

V2X-ViT: 686.8 GFLOPs

CoBEVT: 216 GFLOPs

Yolov8m: 78.9 GFLOPs

VirConv: N/A

3. Estimate power consumption

AI Model	Estimated power consumption (Nvidia A100 for Nvidia HGX)	Estimated power consumption (Nvidia A100 for PCle)
V2X-ViT	28.3 W	17.7 W
CoBEVT	8.9 W	5.6 W
Yolov8	3.3 W	2.0 W
VirConv		

Estimated Power Consumption by Model and Configuration

Resources:

https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet.pdf

(13) NVIDIA A100 vs V100: How do they compare? | LinkedIn

GPUs Lead in Energy Efficiency, DoE Center Says | NVIDIA Blogs

Reducing Energy Consumption of ML Models | Datatonic

YOLOv8 - Ultralytics YOLO Docs