

Online Result Summary

Model: yolov8n

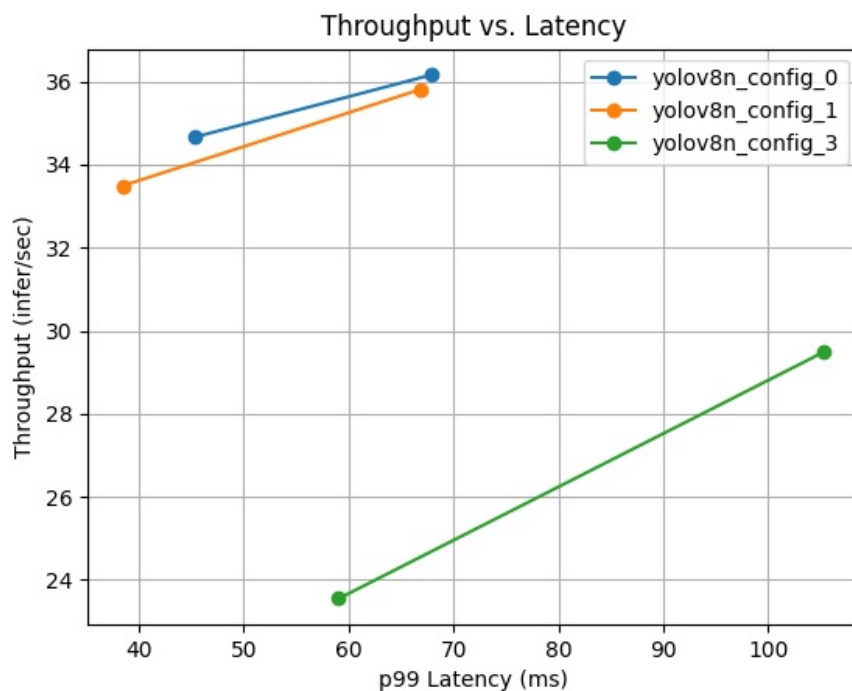
Constraint targets: None

In 8 measurements across 4 configurations, **yolov8n_config_0** provides the best throughput: **36 infer/sec**.

Which provides no gain over the default configuration, under the given constraints.

- **yolov8n_config_0**: 1/CPU model instances with a max batch size of 1 on platform onnxruntime

Curves corresponding to the 3 best model configuration(s) out of a total of 4 are shown in the plots.



Throughput vs. Latency curves for 3 best configurations.

The following table summarizes each configuration at the measurement that optimizes the desired metrics under the given constraints.

Model Config Name	Max Batch Size	Dynamic Batching	Instance Count	p99 Latency (ms)	Throughput (infer/sec)
yolov8n_config_0	1	Enabled	1/CPU	67.873	36.157
yolov8n_config_1	2	Enabled	1/CPU	66.917	35.8149
yolov8n_config_3	2	Enabled	2/CPU	105.347	29.4931