OpenVINO DEEP LEARNING WORKBENCH

← Back to Create Configuration

Configurations   [Create]

Configurations Settings

#	Model	Validation Dataset	Target	Precisions	Creation Time	Latency (ms)	Throughput (FPS)	Accuracy (%)	Status	Actions
1	frozen_inference_graph	Refined30_coco	Machine: Local Workstation Device: CPU	FP32	17/09/20, 11:33	1,215.89	0.83	N/A	⚠	🗑

Accuracy tool failed

[Compare]

Selected Configuration

frozen_inference_graph • Refined30_coco • Local Workstation • CPU

[**Profile**] [Optimize] [Pack]

Select Inference Method

ℹ ● Single Inference

Parallel streams(1-4): ℹ
[1]

Batch size (1-256): ℹ
[1]

ℹ ○ Group Inference

[Execute]

Inference Tips

Combinations of streams and batch sizes appear in real time on the Inference Results plot below and are characterized with throughput and latency values. Inference is executed asynchronously. To profile multiple combinations of parameters in sequence, use group inference.
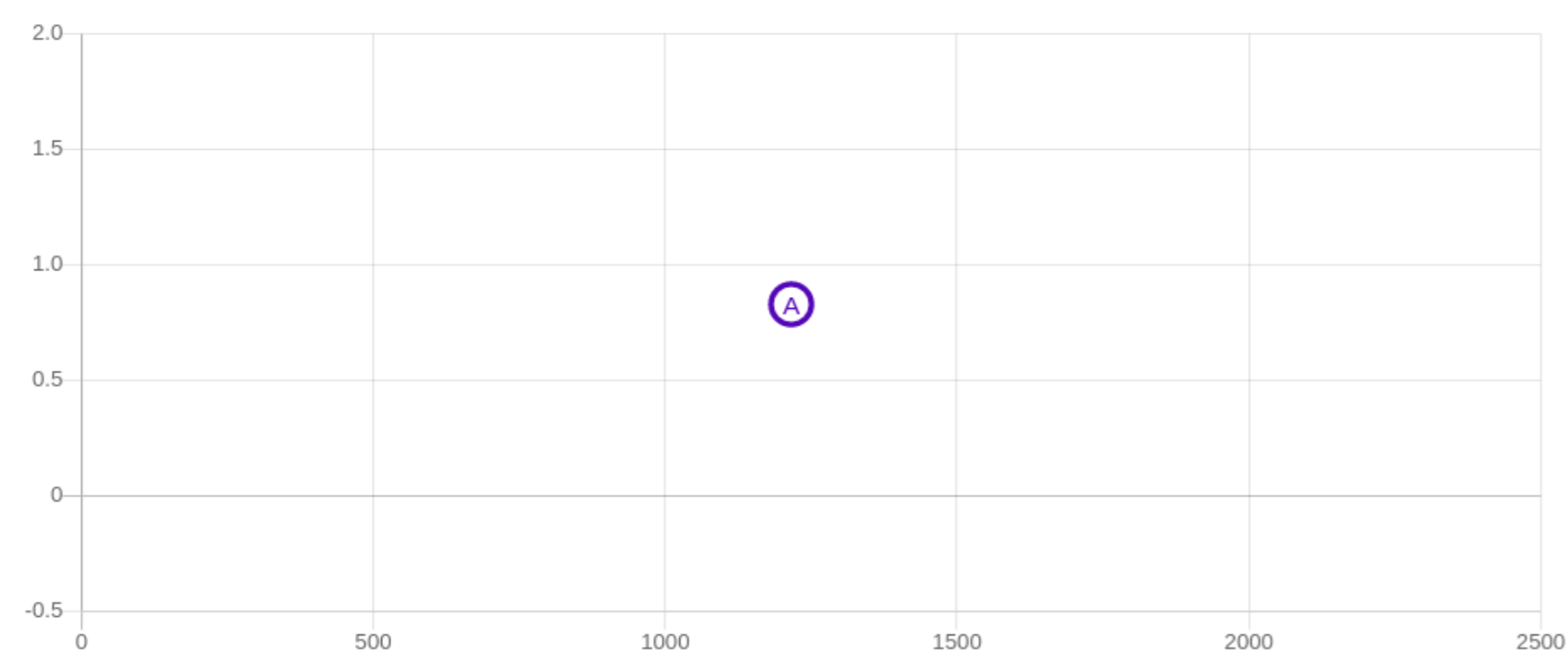
Model Performance Summary

Inference Results

frozen_inference_graph • Refined30_coco • Local Workstation • CPU
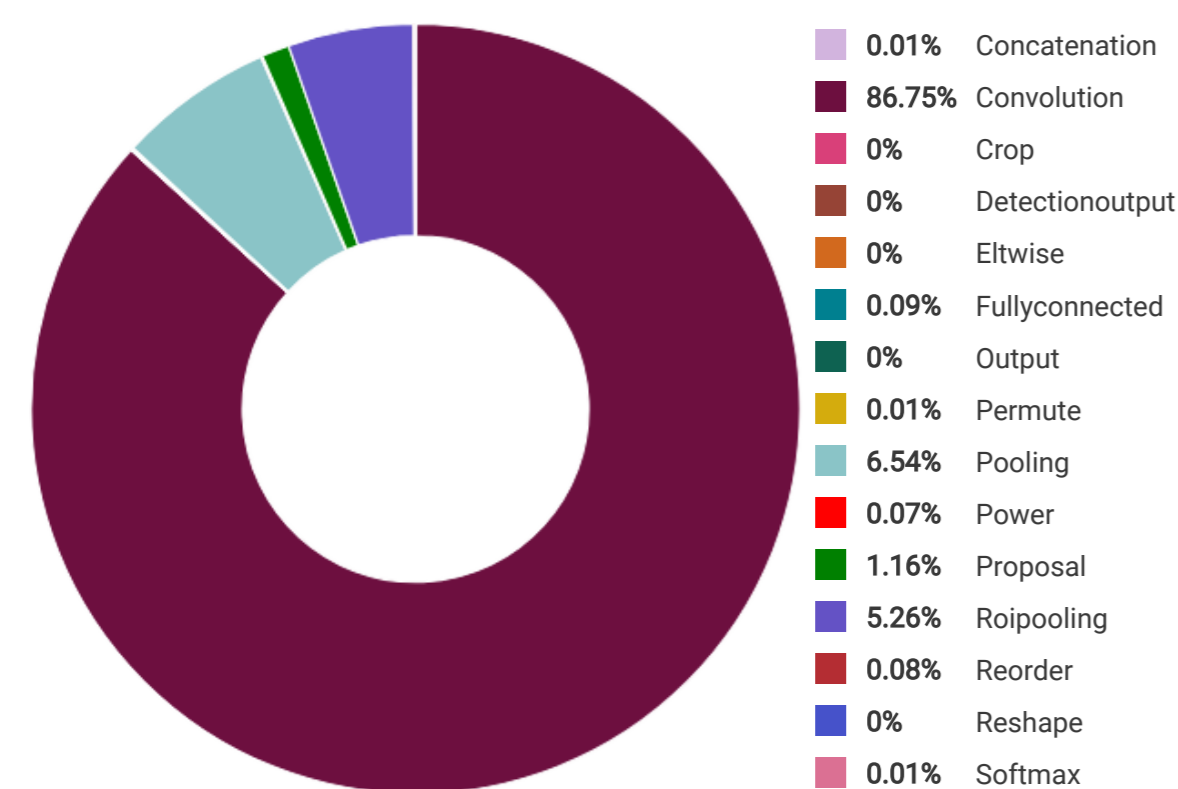
Latency Threshold (0-1000)

Throughput, fps   ⟨⟩ Expand Chart   ⟨⟩ Expand Table

#	Streams ↑	Batch size ↑	Throughput, fps ↑	Latency, ms ↑	Last status	Filter
A	1	1	0.83	1,215.89	✓	☐

Latency, ms

- ● - frozen_inference_graph • Refined30_coco • Local Workstation • CPU

Ⓐ - Selected Point

● - Sweet Spot Point

Execution Time By Layer

frozen_inference_graph • Refined30_coco • Local Workstation • CPU

%	Layer
0.01%	Concatenation
86.75%	Convolution
0%	Crop
0%	Detectionoutput
0%	Eltwise
0.09%	Fullyconnected
0%	Output
0.01%	Permute
6.54%	Pooling
0.07%	Power
1.16%	Proposal
5.26%	Roipooling
0.08%	Reorder
0%	Reshape
0.01%	Softmax

Execution Attributes

frozen_inference_graph • Refined30_coco • Local Workstation • CPU

FPS: 0.83

Latency (ms): 1215.89

Total Execution Time (ms): 21731.81

Batch size: 1

Streams: 1

Layers Table

frozen_inference_graph • Refined30_coco • Local Workstation • CPU

[Visualize original IR] [Visualize runtime graph]

Select Column [▾]   Select Filter [▾]   + Add new filter

[Apply Filter]   Clear Filter

Execution Order	Layer Name	Layer Type	Execution Time, ms	Precision
2	image_tensor	Input	Not Executed	U8
3	image_tensor_U8_FP32_Preprocessor/sub	Reorder	0.628	FP32
4	Preprocessor/sub	Power	0.938	FP32
5	FirstStageFeatureExtractor/InceptionV2/InceptionV2/Conv2d_1a_7x7/separable_conv2d/depthwise	Convolution	7.831	FP32
6	FirstStageFeatureExtractor/InceptionV2/InceptionV2/Conv2d_1a_7x7/BatchNorm/FusedBatchNormV3/variance/Fused_Add_	Convolution	5.552	FP32
7	FirstStageFeatureExtractor/InceptionV2/InceptionV2/MaxPool_2a_3x3/MaxPool	Pooling	5.981	FP32
8	FirstStageFeatureExtractor/InceptionV2/InceptionV2/Conv2d_2b_1x1/BatchNorm/FusedBatchNormV3/variance/Fused_Add_	Convolution	2.995	FP32
9	FirstStageFeatureExtractor/InceptionV2/InceptionV2/Conv2d_2c_3x3/BatchNorm/FusedBatchNormV3/variance/Fused_Add_	Convolution	69.182	FP32
10	FirstStageFeatureExtractor/InceptionV2/InceptionV2/MaxPool_3a_3x3/MaxPool	Pooling	3.372	FP32
11	FirstStageFeatureExtractor/InceptionV2/InceptionV2/Mixed_3b/Branch_0/Conv2d_0a_1x1/BatchNorm/FusedBatchNormV3/variance/Fused_Add_	Convolution	2.099	FP32
12	FirstStageFeatureExtractor/InceptionV2/InceptionV2/Mixed_3b/Branch_1/Conv2d_0a_1x1/BatchNorm/FusedBatchNormV3/variance/Fused_Add_	Convolution	1.921	FP32
13	FirstStageFeatureExtractor/InceptionV2/InceptionV2/Mixed_3b/Branch_1/Conv2d_0b_3x3/BatchNorm/FusedBatchNormV3/variance/Fused_Add_	Convolution	5.557	FP32