Model conversion

1. Jetson Orin YOLO11 (benchmark)

YOLO11 benchmark data comes from the Ultralytics team, which tests models in multiple formats (data is for reference only)

2. Enable optimal performance of the motherboard

Enable Jetson clocks

Enabling Jetson Clocks will ensure that all CPU and GPU cores run at maximum frequency:

```
sudo jetson_clocks
```

3. Model conversion

According to the test parameters of different formats provided by the Ultralytics team, we can find that the inference performance is best when using TensorRT!

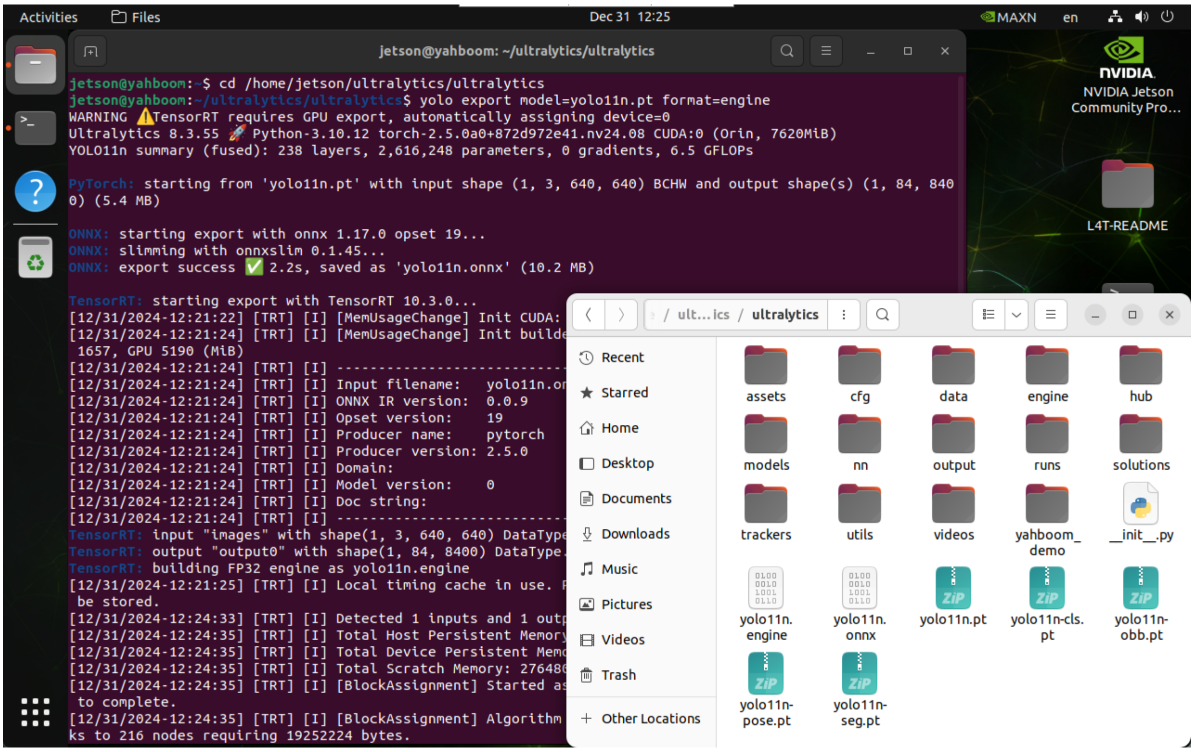
```
When using the export mode of YOLO11 for the first time, some dependencies will
be automatically installed. Just wait for it to be completed automatically!
```

3.1. CLI: pt → onnx → engine

Convert the PyTorch format model to TensorRT: The conversion process will automatically generate an ONNX model

```
cd /home/jetson/ultralytics/ultralytics
```

```
yolo export model=yolo11n.pt format=engine
# yolo export model=yolo11n-seg.pt format=engine
# yolo export model=yolo11n-pose.pt format=engine
# yolo export model=yolo11n-cls.pt format=engine
# yolo export model=yolo11n-obb.pt format=engine
```



```
jetson@yahboom:~$ cd /home/jetson/ultralytics/ultralytics
jetson@yahboom:~/ultralytics/ultralytics$ yolo export model=yolo11n.pt format=engine
WARNING ⚠️TensorRT requires GPU export, automatically assigning device=0
Ultralytics 8.3.55 🚀 Python-3.10.12 torch-2.5.0a0+872d972e41.nv24.08 CUDA:0 (Orin, 7620MiB)
YOLO11n summary (fused): 238 layers, 2,616,248 parameters, 0 gradients, 6.5 GFLOPs

PyTorch: starting from 'yolo11n.pt' with input shape (1, 3, 640, 640) BCHW and output shape(s) (1, 84, 840
0) (5.4 MB)

ONNX: starting export with onnx 1.17.0 opset 19...
ONNX: slimming with onnxslim 0.1.45...
ONNX: export success ✅ 2.2s, saved as 'yolo11n.onnx' (10.2 MB)

TensorRT: starting export with TensorRT 10.3.0...
[12/31/2024-12:21:22] [TRT] [I] [MemUsageChange] Init CUDA:
[12/31/2024-12:21:24] [TRT] [I] [MemUsageChange] Init builde
 1657, GPU 5190 (MiB)
[12/31/2024-12:21:24] [TRT] [I] ----------------------------
[12/31/2024-12:21:24] [TRT] [I] Input filename:   yolo11n.or
[12/31/2024-12:21:24] [TRT] [I] ONNX IR version:  0.0.9
[12/31/2024-12:21:24] [TRT] [I] Opset version:    19
[12/31/2024-12:21:24] [TRT] [I] Producer name:    pytorch
[12/31/2024-12:21:24] [TRT] [I] Producer version: 2.5.0
[12/31/2024-12:21:24] [TRT] [I] Domain:
[12/31/2024-12:21:24] [TRT] [I] Model version:    0
[12/31/2024-12:21:24] [TRT] [I] Doc string:
[12/31/2024-12:21:24] [TRT] [I] ----------------------------
TensorRT: input "images" with shape(1, 3, 640, 640) DataType
TensorRT: output "output0" with shape(1, 84, 8400) DataType.
TensorRT: building FP32 engine as yolo11n.engine
[12/31/2024-12:21:25] [TRT] [I] Local timing cache in use. F
 be stored.
[12/31/2024-12:24:33] [TRT] [I] Detected 1 inputs and 1 outp
[12/31/2024-12:24:35] [TRT] [I] Total Host Persistent Memory
[12/31/2024-12:24:35] [TRT] [I] Total Device Persistent Memo
[12/31/2024-12:24:35] [TRT] [I] Total Scratch Memory: 276480
[12/31/2024-12:24:35] [TRT] [I] [BlockAssignment] Started as
 to complete.
[12/31/2024-12:24:35] [TRT] [I] [BlockAssignment] Algorithm
ks to 216 nodes requiring 19252224 bytes.
```

3.2、Python：pt → onnx → engine

Convert PyTorch format models to TensorRT: The conversion process will automatically generate ONNX models

```
cd /home/jetson/ultralytics/ultralytics/yahboom_demo
```

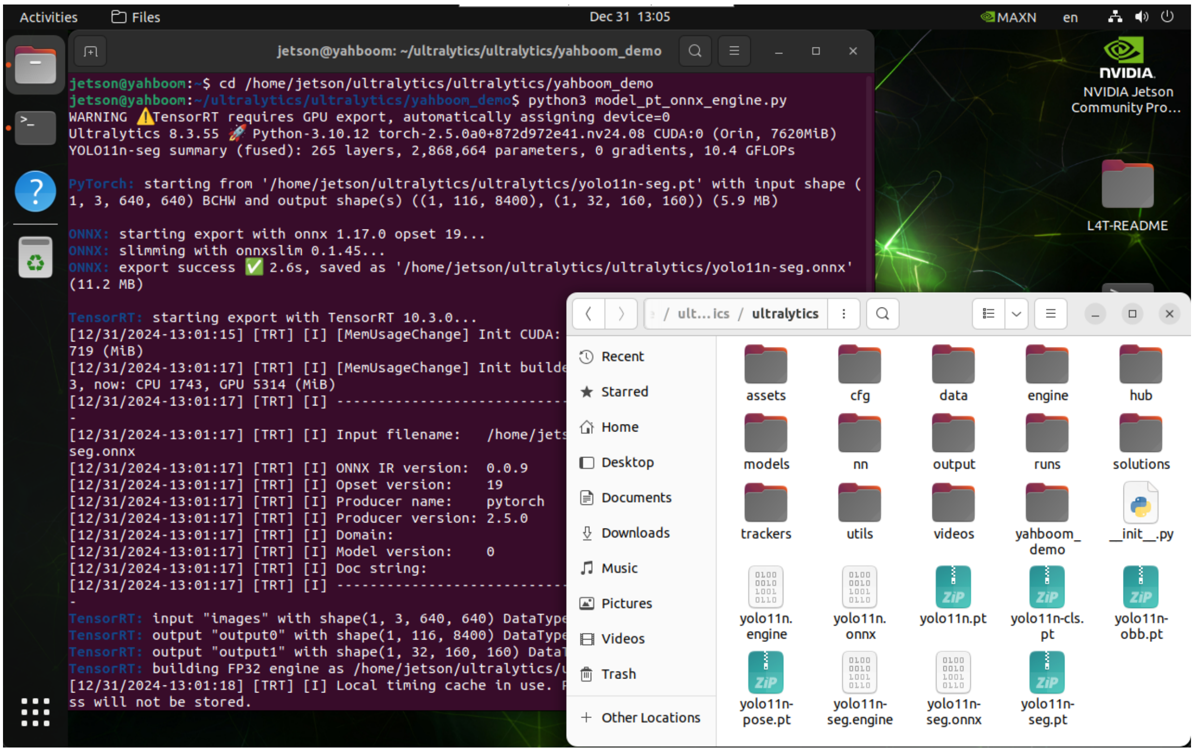
```
python3 model_pt_onnx_engine.py
```

```
from ultralytics import YOLO

# Load a YOLO11n PyTorch model
# model = YOLO("/home/jetson/ultralytics/ultralytics/yolo11n.pt")
model = YOLO("/home/jetson/ultralytics/ultralytics/yolo11n-seg.pt")
# model = YOLO("/home/jetson/ultralytics/ultralytics/yolo11n-pose.pt")
# model = YOLO("/home/jetson/ultralytics/ultralytics/yolo11n-cls.pt")
# model = YOLO("/home/jetson/ultralytics/ultralytics/yolo11n-obb.pt")

# Export the model to TensorRT
model.export(format="engine")
```

Note: The model file generated by the conversion is located in the converted model file location



```
jetson@yahboom:~$ cd /home/jetson/ultralytics/ultralytics/yahboom_demo
jetson@yahboom:~/ultralytics/ultralytics/yahboom_demo$ python3 model_pt_onnx_engine.py
WARNING ⚠️TensorRT requires GPU export, automatically assigning device=0
Ultralytics 8.3.55 🚀 Python-3.10.12 torch-2.5.0a0+872d972e41.nv24.08 CUDA:0 (Orin, 7620MiB)
YOLO11n-seg summary (fused): 265 layers, 2,868,664 parameters, 0 gradients, 10.4 GFLOPs

PyTorch: starting from '/home/jetson/ultralytics/ultralytics/yolo11n-seg.pt' with input shape (
1, 3, 640, 640) BCHW and output shape(s) ((1, 116, 8400), (1, 32, 160, 160)) (5.9 MB)

ONNX: starting export with onnx 1.17.0 opset 19...
ONNX: slimming with onnxslim 0.1.45...
ONNX: export success ✅ 2.6s, saved as '/home/jetson/ultralytics/ultralytics/yolo11n-seg.onnx'
(11.2 MB)

TensorRT: starting export with TensorRT 10.3.0...
[12/31/2024-13:01:15] [TRT] [I] [MemUsageChange] Init CUDA:
719 (MiB)
[12/31/2024-13:01:17] [TRT] [I] [MemUsageChange] Init builde
3, now: CPU 1743, GPU 5314 (MiB)
[12/31/2024-13:01:17] [TRT] [I] -----------------------------
-
[12/31/2024-13:01:17] [TRT] [I] Input filename:   /home/jets
seg.onnx
[12/31/2024-13:01:17] [TRT] [I] ONNX IR version:  0.0.9
[12/31/2024-13:01:17] [TRT] [I] Opset version:    19
[12/31/2024-13:01:17] [TRT] [I] Producer name:    pytorch
[12/31/2024-13:01:17] [TRT] [I] Producer version: 2.5.0
[12/31/2024-13:01:17] [TRT] [I] Domain:
[12/31/2024-13:01:17] [TRT] [I] Model version:    0
[12/31/2024-13:01:17] [TRT] [I] Doc string:
[12/31/2024-13:01:17] [TRT] [I] -----------------------------
-
TensorRT: input "images" with shape(1, 3, 640, 640) DataType
TensorRT: output "output0" with shape(1, 116, 8400) DataType
TensorRT: output "output1" with shape(1, 32, 160, 160) DataT
TensorRT: building FP32 engine as /home/jetson/ultralytics/u
[12/31/2024-13:01:18] [TRT] [I] Local timing cache in use. P
ss will not be stored.
```

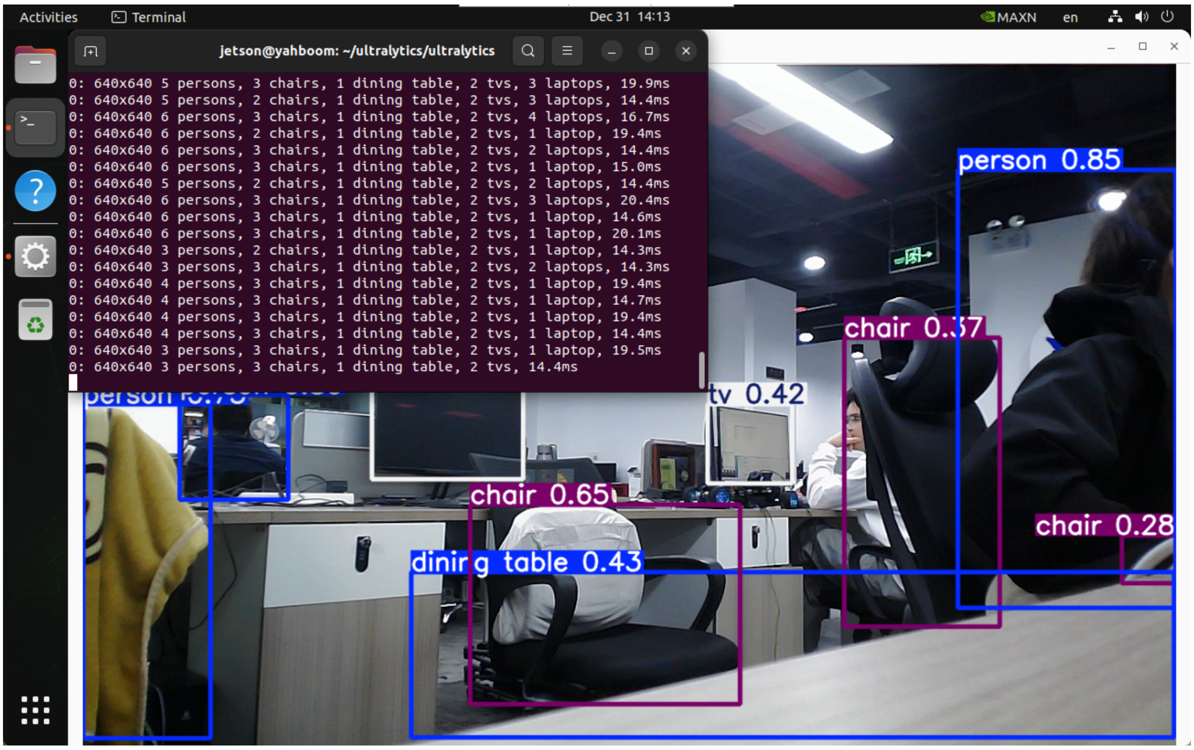
4. Model prediction

CLI usage

CLI currently only supports calling USB cameras. Nuwa camera users can directly modify the previous python code to call onnx and engine models!

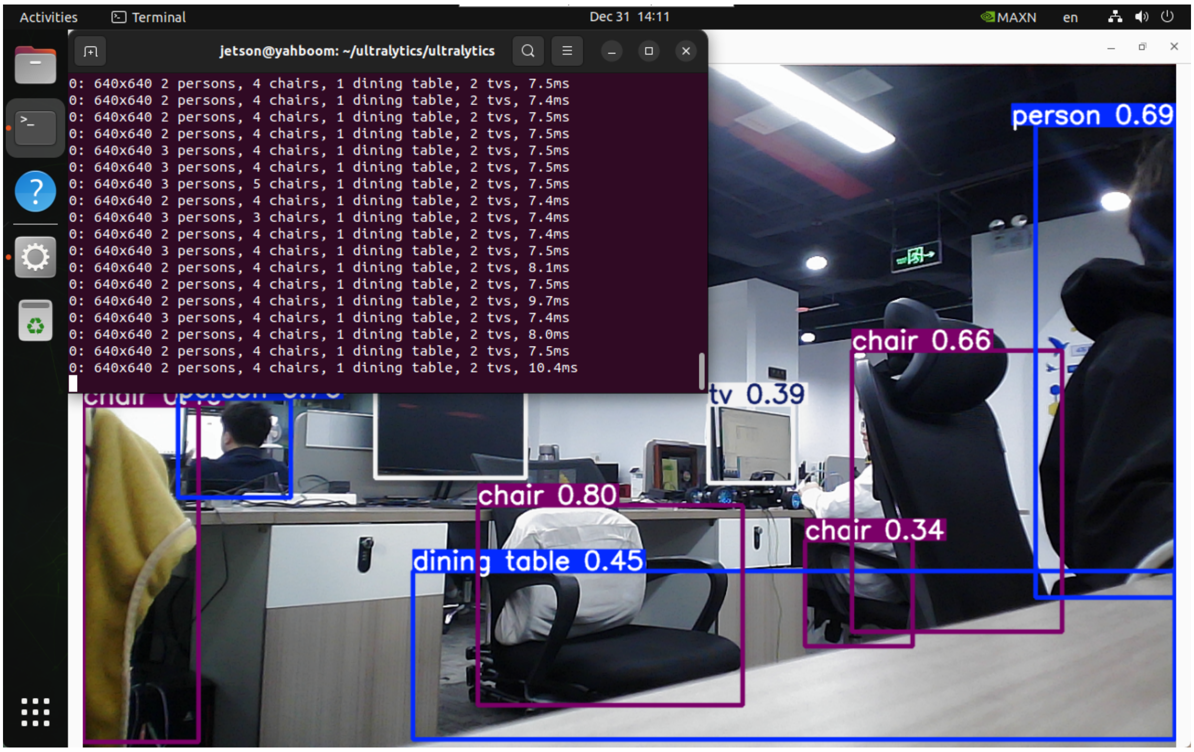
```
cd /home/jetson/ultralytics/ultralytics
```

```
yolo predict model=yolo11n.onnx source=0 save=False show
```



```
0: 640x640 5 persons, 3 chairs, 1 dining table, 2 tvs, 3 laptops, 19.9ms
0: 640x640 5 persons, 2 chairs, 1 dining table, 2 tvs, 3 laptops, 14.4ms
0: 640x640 6 persons, 3 chairs, 1 dining table, 2 tvs, 4 laptops, 16.7ms
0: 640x640 6 persons, 2 chairs, 1 dining table, 2 tvs, 1 laptop, 19.4ms
0: 640x640 6 persons, 3 chairs, 1 dining table, 2 tvs, 2 laptops, 14.4ms
0: 640x640 6 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 15.0ms
0: 640x640 5 persons, 2 chairs, 1 dining table, 2 tvs, 2 laptops, 14.4ms
0: 640x640 6 persons, 3 chairs, 1 dining table, 2 tvs, 3 laptops, 20.4ms
0: 640x640 6 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 14.6ms
0: 640x640 6 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 20.1ms
0: 640x640 3 persons, 2 chairs, 1 dining table, 2 tvs, 1 laptop, 14.3ms
0: 640x640 3 persons, 3 chairs, 1 dining table, 2 tvs, 2 laptops, 14.3ms
0: 640x640 4 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 19.4ms
0: 640x640 4 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 14.7ms
0: 640x640 4 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 19.4ms
0: 640x640 4 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 14.4ms
0: 640x640 3 persons, 3 chairs, 1 dining table, 2 tvs, 1 laptop, 19.5ms
0: 640x640 3 persons, 3 chairs, 1 dining table, 2 tvs, 14.4ms
```

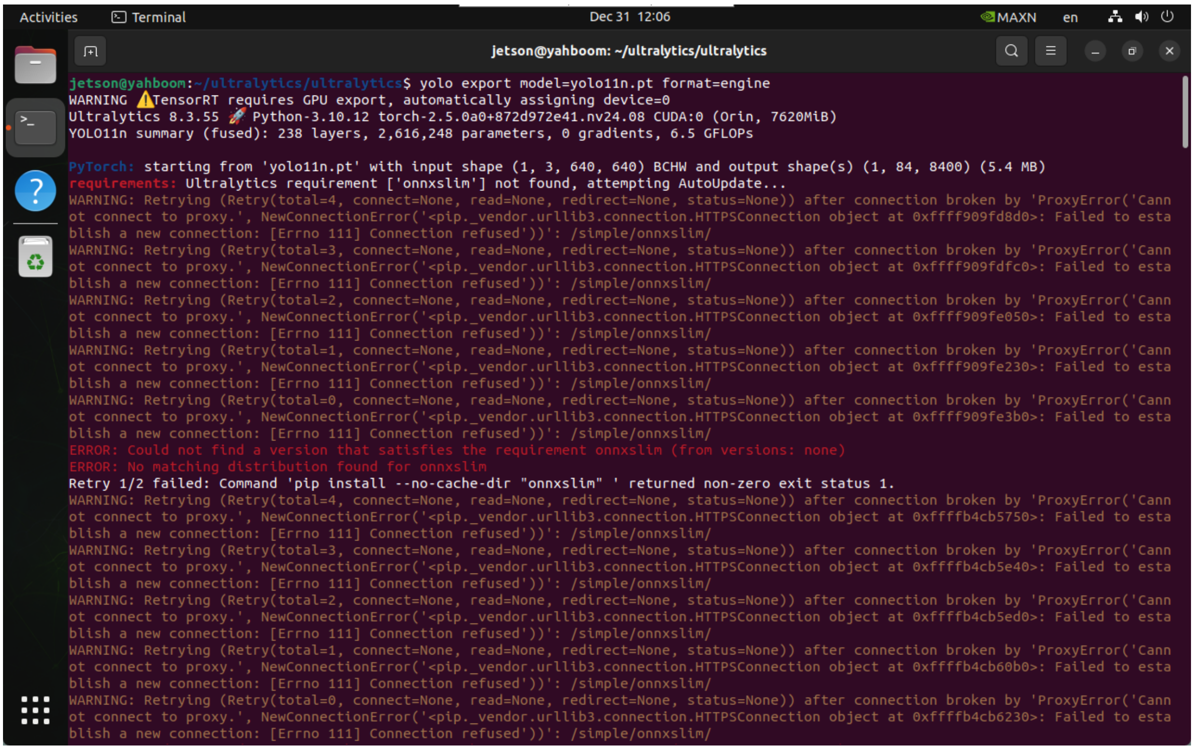
```
yolo predict model=yolo11n.engine source=0 save=False show
```



```
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.4ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 3 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 3 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 3 persons, 5 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.4ms
0: 640x640 3 persons, 3 chairs, 1 dining table, 2 tvs, 7.4ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.4ms
0: 640x640 3 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 8.1ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 9.7ms
0: 640x640 3 persons, 4 chairs, 1 dining table, 2 tvs, 7.4ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 8.0ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 7.5ms
0: 640x640 2 persons, 4 chairs, 1 dining table, 2 tvs, 10.4ms
```

Frequently asked questions

ERROR: onnxslim

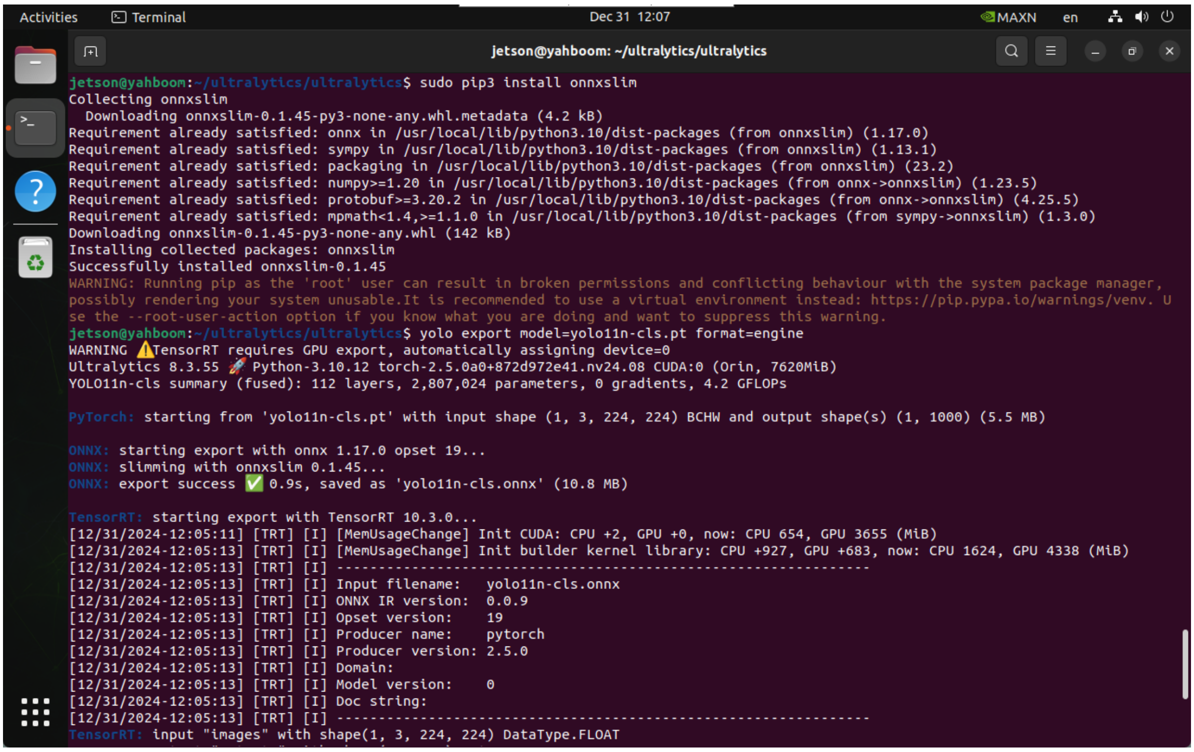


```
jetson@yahboom:~/ultralytics/ultralytics$ yolo export model=yolo11n.pt format=engine
WARNING ⚠️TensorRT requires GPU export, automatically assigning device=0
Ultralytics 8.3.55 🚀 Python-3.10.12 torch-2.5.0a0+872d972e41.nv24.08 CUDA:0 (Orin, 7620MiB)
YOLO11n summary (fused): 238 layers, 2,616,248 parameters, 0 gradients, 6.5 GFLOPs

PyTorch: starting from 'yolo11n.pt' with input shape (1, 3, 640, 640) BCHW and output shape(s) (1, 84, 8400) (5.4 MB)
requirements: Ultralytics requirement ['onnxslim'] not found, attempting AutoUpdate...
WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffff909fd8d0>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=3, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffff909fdfc0>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=2, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffff909fe050>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffff909fe230>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=0, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffff909fe3b0>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
ERROR: Could not find a version that satisfies the requirement onnxslim (from versions: none)
ERROR: No matching distribution found for onnxslim
Retry 1/2 failed: Command 'pip install --no-cache-dir "onnxslim" ' returned non-zero exit status 1.
WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffffb4cb5750>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=3, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffffb4cb5e40>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=2, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffffb4cb5ed0>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffffb4cb60b0>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
WARNING: Retrying (Retry(total=0, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProxyError('Cann
ot connect to proxy.', NewConnectionError('<pip._vendor.urllib3.connection.HTTPSConnection object at 0xffffb4cb6230>: Failed to esta
blish a new connection: [Errno 111] Connection refused'))': /simple/onnxslim/
```

Solution: Enter the onnxslim installation command in the terminal

```
sudo pip3 install onnxslim
```



```
jetson@yahboom:~/ultralytics/ultralytics$ sudo pip3 install onnxslim
Collecting onnxslim
  Downloading onnxslim-0.1.45-py3-none-any.whl.metadata (4.2 kB)
Requirement already satisfied: onnx in /usr/local/lib/python3.10/dist-packages (from onnxslim) (1.17.0)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from onnxslim) (1.13.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from onnxslim) (23.2)
Requirement already satisfied: numpy>=1.20 in /usr/local/lib/python3.10/dist-packages (from onnx->onnxslim) (1.23.5)
Requirement already satisfied: protobuf>=3.20.2 in /usr/local/lib/python3.10/dist-packages (from onnx->onnxslim) (4.25.5)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from sympy->onnxslim) (1.3.0)
Downloading onnxslim-0.1.45-py3-none-any.whl (142 kB)
Installing collected packages: onnxslim
Successfully installed onnxslim-0.1.45
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager,
possibly rendering your system unusable.It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv. U
se the --root-user-action option if you know what you are doing and want to suppress this warning.
jetson@yahboom:~/ultralytics/ultralytics$ yolo export model=yolo11n-cls.pt format=engine
WARNING ⚠️TensorRT requires GPU export, automatically assigning device=0
Ultralytics 8.3.55 🚀 Python-3.10.12 torch-2.5.0a0+872d972e41.nv24.08 CUDA:0 (Orin, 7620MiB)
YOLO11n-cls summary (fused): 112 layers, 2,807,024 parameters, 0 gradients, 4.2 GFLOPs

PyTorch: starting from 'yolo11n-cls.pt' with input shape (1, 3, 224, 224) BCHW and output shape(s) (1, 1000) (5.5 MB)

ONNX: starting export with onnx 1.17.0 opset 19...
ONNX: slimming with onnxslim 0.1.45...
ONNX: export success ✅ 0.9s, saved as 'yolo11n-cls.onnx' (10.8 MB)

TensorRT: starting export with TensorRT 10.3.0...
[12/31/2024-12:05:11] [TRT] [I] [MemUsageChange] Init CUDA: CPU +2, GPU +0, now: CPU 654, GPU 3655 (MiB)
[12/31/2024-12:05:13] [TRT] [I] [MemUsageChange] Init builder kernel library: CPU +927, GPU +683, now: CPU 1624, GPU 4338 (MiB)
[12/31/2024-12:05:13] [TRT] [I] ----------------------------------------------------------------
[12/31/2024-12:05:13] [TRT] [I] Input filename:   yolo11n-cls.onnx
[12/31/2024-12:05:13] [TRT] [I] ONNX IR version:  0.0.9
[12/31/2024-12:05:13] [TRT] [I] Opset version:    19
[12/31/2024-12:05:13] [TRT] [I] Producer name:    pytorch
[12/31/2024-12:05:13] [TRT] [I] Producer version: 2.5.0
[12/31/2024-12:05:13] [TRT] [I] Domain:
[12/31/2024-12:05:13] [TRT] [I] Model version:    0
[12/31/2024-12:05:13] [TRT] [I] Doc string:
[12/31/2024-12:05:13] [TRT] [I] ----------------------------------------------------------------
TensorRT: input "images" with shape(1, 3, 224, 224) DataType.FLOAT
```

References

https://docs.ultralytics.com/guides/nvidia-jetson/

https://docs.ultralytics.com/integrations/tensorrt/