```
!pip install --no-deps bitsandbytes accelerate xformers==0.0.29.post3
peft trl triton cut_cross_entropy unsloth_zoo
!pip install sentencepiece protobuf datasets huggingface_hub
hf_transfer
!pip install --no-deps unsloth

Collecting bitsandbytes
  Downloading bitsandbytes-0.48.1-py3-none-
manylinux_2_24_x86_64.whl.metadata (10 kB)
Requirement already satisfied: accelerate in
/usr/local/lib/python3.12/dist-packages (1.10.1)
Collecting xformers==0.0.29.post3
  Downloading xformers-0.0.29.post3-cp312-cp312-
manylinux_2_28_x86_64.whl.metadata (1.2 kB)
Requirement already satisfied: peft in /usr/local/lib/python3.12/dist-
packages (0.17.1)
Collecting trl
  Downloading trl-0.23.1-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: triton in
/usr/local/lib/python3.12/dist-packages (3.4.0)
Collecting cut_cross_entropy
  Downloading cut_cross_entropy-25.1.1-py3-none-any.whl.metadata (9.3
kB)
Collecting unsloth_zoo
  Downloading unsloth_zoo-2025.9.14-py3-none-any.whl.metadata (31 kB)
Downloading xformers-0.0.29.post3-cp312-cp312-
manylinux_2_28_x86_64.whl (43.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43.4/43.4 MB 21.6 MB/s eta
0:00:00
anylinux_2_24_x86_64.whl (60.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 60.1/60.1 MB 18.4 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 564.6/564.6 kB 37.9 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 256.5/256.5 kB 22.5 MB/s eta
0:00:00
ers, unsloth_zoo, trl, cut_cross_entropy, bitsandbytes
Successfully installed bitsandbytes-0.48.1 cut_cross_entropy-25.1.1
trl-0.23.1 unsloth_zoo-2025.9.14 xformers-0.0.29.post3
Requirement already satisfied: sentencepiece in
/usr/local/lib/python3.12/dist-packages (0.2.1)
Requirement already satisfied: protobuf in
/usr/local/lib/python3.12/dist-packages (5.29.5)
Requirement already satisfied: datasets in
/usr/local/lib/python3.12/dist-packages (4.0.0)
Requirement already satisfied: huggingface_hub in
/usr/local/lib/python3.12/dist-packages (0.35.3)
Requirement already satisfied: hf_transfer in
/usr/local/lib/python3.12/dist-packages (0.1.9)
Requirement already satisfied: filelock in
```

```
/usr/local/lib/python3.12/dist-packages (from datasets) (3.19.1)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.12/dist-packages (from datasets) (2.0.2)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.12/dist-packages (from datasets) (18.1.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in
/usr/local/lib/python3.12/dist-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in
/usr/local/lib/python3.12/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: requests>=2.32.2 in
/usr/local/lib/python3.12/dist-packages (from datasets) (2.32.4)
Requirement already satisfied: tqdm>=4.66.3 in
/usr/local/lib/python3.12/dist-packages (from datasets) (4.67.1)
Requirement already satisfied: xxhash in
/usr/local/lib/python3.12/dist-packages (from datasets) (3.5.0)
Requirement already satisfied: multiprocess<0.70.17 in
/usr/local/lib/python3.12/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2025.3.0,>=2023.1.0 in
/usr/local/lib/python3.12/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2025.3.0)
Requirement already satisfied: packaging in
/usr/local/lib/python3.12/dist-packages (from datasets) (25.0)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.12/dist-packages (from datasets) (6.0.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.12/dist-packages (from huggingface_hub)
(4.15.0)
Requirement already satisfied: hf-xet<2.0.0,>=1.1.3 in
/usr/local/lib/python3.12/dist-packages (from huggingface_hub)
(1.1.10)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in
/usr/local/lib/python3.12/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (3.12.15)
Requirement already satisfied: charset_normalizer<4,>=2 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.32.2-
>datasets) (3.4.3)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.32.2-
>datasets) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.32.2-
>datasets) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.32.2-
>datasets) (2025.8.3)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.12/dist-packages (from pandas->datasets)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
```

```
/usr/local/lib/python3.12/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.12/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: aiohappyeyeballs>=2.5.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.4.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.4.0)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.7.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (6.6.4)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (0.3.2)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.20.1)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Collecting unsloth
  Downloading unsloth-2025.9.11-py3-none-any.whl.metadata (55 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 55.1/55.1 kB 2.3 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 317.5/317.5 kB 8.8 MB/s eta
0:00:00

from unsloth import FastVisionModel
import torch

🦥 Unsloth: Will patch your computer to enable 2x faster free
finetuning.

WARNING:xformers:WARNING[XFORMERS]: xFormers can't load C++/CUDA
extensions. xFormers was built for:
    PyTorch 2.6.0+cu124 with CUDA 1204 (you have 2.8.0+cu126)
    Python  3.12.9 (you have 3.12.11)
  Please reinstall xformers (see
https://github.com/facebookresearch/xformers#installing-xformers)
  Memory-efficient attention, SwiGLU, sparse and more won't be
```

```
available.
  Set XFORMERS_MORE_DETAILS=1 for more details

========
Switching to PyTorch attention since your Xformers is broken.
========

Unsloth: Xformers was not installed correctly.
Please install xformers separately first.
Then confirm if it's correctly installed by running:
python -m xformers.info

Longer error message:
xFormers can't load C++/CUDA extensions. xFormers was built for:
    PyTorch 2.6.0+cu124 with CUDA 1204 (you have 2.8.0+cu126)
    Python  3.12.9 (you have 3.12.11)
  Please reinstall xformers (see
https://github.com/facebookresearch/xformers#installing-xformers)
  Memory-efficient attention, SwiGLU, sparse and more won't be
available.
 Unsloth Zoo will now patch everything to make training faster!
```

```
fourbit_model = [
    "unsloth/llama-3.2-11B-Vision-Instruct-bnb-4bit",
    "unsloth/Qwen2-VL-7B-Instruct-bnb-4bit"
]

model, tokenizer = FastVisionModel.from_pretrained(
    "unsloth/Qwen2-VL-7B-Instruct",
    load_in_4bit=True,
    use_gradient_checkpointing="unsloth"
)
```

```
==((====))==  Unsloth 2025.9.11: Fast Qwen2_Vl patching. Transformers:
4.56.2.
   \\   /|    Tesla T4. Num GPUs = 1. Max memory: 14.741 GB. Platform:
Linux.
O^O/ \_/ \    Torch: 2.8.0+cu126. CUDA: 7.5. CUDA Toolkit: 12.6.
Triton: 3.4.0
\        /    Bfloat16 = FALSE. FA [Xformers = None. FA2 = False]
 "-____-"     Free license: http://github.com/unslothai/unsloth
Unsloth: Fast downloading is enabled - ignore downloading bars which
are red colored!
```


```
{"model_id":"cfaf4d1f143c42d1b924f018b4a45423","version_major":2,"vers
ion_minor":0}

{"model_id":"2ea94bddc48a4a9ebb83b3615b50d894","version_major":2,"vers
ion_minor":0}
```

{"model_id":"2a594671d56749609f50d9c818ba2734","version_major":2,"version_minor":0}

```
The image processor of type `Qwen2VLImageProcessor` is now loaded as a fast processor by default, even if the model checkpoint was saved with a slow processor. This is a breaking change and may produce slightly different outputs. To continue using the slow processor, instantiate this class with `use_fast=False`. Note that this behavior will be extended to all models in a future release.
```

{"model_id":"dc4c0354cc0b408c9b037f9be0895363","version_major":2,"version_minor":0}

{"model_id":"0b532ca8c14044edbfcc23e106a5714f","version_major":2,"version_minor":0}

{"model_id":"70837ee3a5cc457eb0a3c5f5cf444ac1","version_major":2,"version_minor":0}

{"model_id":"8d2d716a610d431a93e12e8cae25d0e8","version_major":2,"version_minor":0}

{"model_id":"9424c46a3cfa428db30a03a41e64f597","version_major":2,"version_minor":0}

{"model_id":"fe0c460ee6ee4a8b938709e6400ae5bb","version_major":2,"version_minor":0}

{"model_id":"298687a3ae3b4281930d58aa198a33ab","version_major":2,"version_minor":0}

```
model = FastVisionModel.get_peft_model(
    model,
    finetune_vision_layers=True,
    finetune_language_layers=True,
    finetune_attention_modules=True,
    finetune_mlp_modules=True,

    r=16,
    lora_alpha=16,
    lora_dropout=0,
    bias="none",
    random_state = 3047,
    use_rslora=False,
    loftq_rslora=None


)
```

```
Unsloth: Making `model.base_model.model.model.visual` require gradients
```

```
from datasets import load_dataset
dataset = load_dataset("unsloth/Latex_OCR", split="train")
```


```
{"model_id":"442609ccb4cf43b0accc87aa38dc5bfd","version_major":2,"vers
ion_minor":0}

{"model_id":"311729c8d42f40f4b0d60182f5fab4c4","version_major":2,"vers
ion_minor":0}

{"model_id":"d8bb41783ba444218cf5da2e0fb0752e","version_major":2,"vers
ion_minor":0}

{"model_id":"753a2cd7848641e89766e6c38b707a36","version_major":2,"vers
ion_minor":0}

{"model_id":"39938347404944329b9747f89aaba598","version_major":2,"vers
ion_minor":0}
```


```
dataset
```

```
Dataset({
    features: ['image', 'text'],
    num_rows: 68686
})
```

```
dataset[0]
```

```
{'image': <PIL.PngImagePlugin.PngImageFile image mode=RGB
size=160x40>,
 'text': '{ \\frac { N } { M } } \\in { \\bf Z } , { \\frac { M }
{ P } } \\in { \\bf Z } , { \\frac { P } { Q } } \\in { \\bf Z }'}
```

```
dataset[0]["image"]
```

$$\frac{N}{M} \in \mathbf{Z}, \frac{M}{P} \in \mathbf{Z}, \frac{P}{Q} \in \mathbf{Z}$$

```
dataset[1]["image"]
```

$$D_{\mu}^{\alpha\beta} \bar{A}_{\mu}^{\alpha\beta} = 0,$$

```
dataset[1]["text"]
```


```
{"type":"string"}
```


```
instruction = "Write a Latex  representation for this image."
```

```
def convert_to_conversation(sample):
  conversation = [
      {"role": "user",
       "content":[
           {"type": "text", "text": instruction},
           {"type": "image", "image": sample["image"]}
       ]
       },
            {"role": "assistant",
       "content":[
           {"type": "text", "text": sample["text"]}
       ]
      }
  ]
  return{"messages": conversation}

convert_to_conversation(dataset[0])
```

```
{'messages': [{'role': 'user',
   'content': [{'type': 'text',
     'text': 'Write a Latex  representation for this image.'},
    {'type': 'image',
     'image': <PIL.PngImagePlugin.PngImageFile image mode=RGB
size=160x40>}]},
  {'role': 'assistant',
   'content': [{'type': 'text',
     'text': '{ \\frac { N } { M } } \\in { \\bf Z } , { \\frac { M }
{ P } } \\in { \\bf Z } , { \\frac { P } { Q } } \\in { \\bf Z }'}]}]}
```

```
converted_dataset = [convert_to_conversation(sample) for sample in
dataset]

converted_dataset[1]
```

```
{'messages': [{'role': 'user',
   'content': [{'type': 'text',
     'text': 'Write a Latex  representation for this image.'},
    {'type': 'image',
     'image': <PIL.PngImagePlugin.PngImageFile image mode=RGB
size=120x50>}]},
  {'role': 'assistant',
   'content': [{'type': 'text',
     'text': 'D _ { \\mu } ^ { \\alpha \\beta } \\bar { A } _ { \\mu }
^ { \\alpha \\beta } = 0 ,'}]}]}
```

```
FastVisionModel.for_inference(model)
```

```
PeftModelForCausalLM(
  (base_model): LoraModel(
    (model): Qwen2VLForConditionalGeneration(
      (model): Qwen2VLModel(
```

```
        (visual): Qwen2VisionTransformerPretrainedModel(
          (patch_embed): PatchEmbed(
            (proj): Conv3d(3, 1280, kernel_size=(2, 14, 14),
stride=(2, 14, 14), bias=False)
          )
          (rotary_pos_emb): VisionRotaryEmbedding()
          (blocks): ModuleList(
            (0-18): 19 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
```

```
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (19): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
```

```
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
```

```
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (20-21): 2 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
```

```
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
```

```
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (22): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
```

```
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (23-28): 6 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
```

```
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
```

```
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (29): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
```

```
                )
                (lora_embedding_A): ParameterDict()
                (lora_embedding_B): ParameterDict()
                (lora_magnitude_vector): ModuleDict()
              )
              (proj): lora.Linear(
                (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                (lora_dropout): ModuleDict(
                  (default): Identity()
                )
                (lora_A): ModuleDict(
                  (default): Linear(in_features=1280,
out_features=16, bias=False)
                )
                (lora_B): ModuleDict(
                  (default): Linear(in_features=16,
out_features=1280, bias=False)
                )
                (lora_embedding_A): ParameterDict()
                (lora_embedding_B): ParameterDict()
                (lora_magnitude_vector): ModuleDict()
              )
            )
            (mlp): VisionMlp(
              (fc1): lora.Linear4bit(
                (base_layer): Linear4bit(in_features=1280,
out_features=5120, bias=True)
                (lora_dropout): ModuleDict(
                  (default): Identity()
                )
                (lora_A): ModuleDict(
                  (default): Linear(in_features=1280,
out_features=16, bias=False)
                )
                (lora_B): ModuleDict(
                  (default): Linear(in_features=16,
out_features=5120, bias=False)
                )
                (lora_embedding_A): ParameterDict()
                (lora_embedding_B): ParameterDict()
                (lora_magnitude_vector): ModuleDict()
              )
              (act): QuickGELUActivation()
              (fc2): lora.Linear4bit(
                (base_layer): Linear4bit(in_features=5120,
out_features=1280, bias=True)
                (lora_dropout): ModuleDict(
                  (default): Identity()
                )
```

```
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (30-31): 2 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
```

```
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
          )
```

```
          (merger): PatchMerger(
            (ln_q): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
            (mlp): Sequential(
              (0): Linear(in_features=5120, out_features=5120,
bias=True)
              (1): GELU(approximate='none')
              (2): Linear(in_features=5120, out_features=3584,
bias=True)
            )
          )
        )
        (language_model): Qwen2VLTextModel(
          (embed_tokens): Embedding(152064, 3584, padding_idx=151654)
          (layers): ModuleList(
            (0-27): 28 x Qwen2VLDecoderLayer(
              (self_attn): Qwen2VLAttention(
                (q_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=3584, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (k_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=512, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=512, bias=False)
                  )
```

```
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (v_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=512, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=512, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (o_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=3584, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (rotary_emb): Qwen2VLRotaryEmbedding()
              )
              (mlp): Qwen2MLP(
                (gate_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=18944, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
```

```
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=18944, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (up_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=18944, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=18944, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (down_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=18944,
out_features=3584, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=18944,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act_fn): SiLU()
```

```
              )
              (input_layernorm): Qwen2RMSNorm((3584,), eps=1e-06)
              (post_attention_layernorm): Qwen2RMSNorm((3584,),
eps=1e-06)
            )
          )
          (norm): Qwen2RMSNorm((3584,), eps=1e-06)
          (rotary_emb): Qwen2VLRotaryEmbedding()
        )
      )
      (lm_head): Linear(in_features=3584, out_features=152064,
bias=False)
    )
  )
)

image = dataset[1] ["image"]
messages = [
    {
        "role": "user",
        "content": [
            {"type": "text", "text": instruction},
            {"type": "image", "image": image}
        ]
    }
]

input_text = tokenizer.apply_chat_template(messages,
add_generation_prompt=True)
input = tokenizer(
    image, input_text,
    add_special_tokens = False,
    return_tensors = "pt",

).to("cuda")

from transformers import TextStreamer
text_streamer = TextStreamer(tokenizer, skip_prompt=True)
_ = model.generate(**input, streamer= text_streamer, max_new_tokens =
128, use_cache=True, temperature=1.5, min_p=0.1)

The Latex representation for the image is:

\[
D_\mu \tilde{A}^\mu = 0
\]<|im_end|>

image
```

$$D_{\mu}^{\alpha\beta}\bar{A}_{\mu}^{\alpha\beta} = 0,$$

```
from unsloth import is_bf16_supported
from unsloth.trainer import UnslothVisionDataCollator
from trl import SFTTrainer, SFTConfig

FastVisionModel.for_inference(model)
```

```
PeftModelForCausalLM(
  (base_model): LoraModel(
    (model): Qwen2VLForConditionalGeneration(
      (model): Qwen2VLModel(
        (visual): Qwen2VisionTransformerPretrainedModel(
          (patch_embed): PatchEmbed(
            (proj): Conv3d(3, 1280, kernel_size=(2, 14, 14),
stride=(2, 14, 14), bias=False)
          )
          (rotary_pos_emb): VisionRotaryEmbedding()
          (blocks): ModuleList(
            (0-18): 19 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
```

```
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
```

```
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (19): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
```

```
              (mlp): VisionMlp(
                (fc1): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (20-21): 2 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
```

```
                (lora_dropout): ModuleDict(
                  (default): Identity()
                )
                (lora_A): ModuleDict(
                  (default): Linear(in_features=1280,
out_features=16, bias=False)
                )
                (lora_B): ModuleDict(
                  (default): Linear(in_features=16,
out_features=3840, bias=False)
                )
                (lora_embedding_A): ParameterDict()
                (lora_embedding_B): ParameterDict()
                (lora_magnitude_vector): ModuleDict()
              )
              (proj): lora.Linear(
                (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                (lora_dropout): ModuleDict(
                  (default): Identity()
                )
                (lora_A): ModuleDict(
                  (default): Linear(in_features=1280,
out_features=16, bias=False)
                )
                (lora_B): ModuleDict(
                  (default): Linear(in_features=16,
out_features=1280, bias=False)
                )
                (lora_embedding_A): ParameterDict()
                (lora_embedding_B): ParameterDict()
                (lora_magnitude_vector): ModuleDict()
              )
            )
            (mlp): VisionMlp(
              (fc1): lora.Linear(
                (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                (lora_dropout): ModuleDict(
                  (default): Identity()
                )
                (lora_A): ModuleDict(
                  (default): Linear(in_features=1280,
out_features=16, bias=False)
                )
                (lora_B): ModuleDict(
                  (default): Linear(in_features=16,
out_features=5120, bias=False)
                )
```

```
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (22): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
```

```
                (proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
```

```
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (23-28): 6 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
```

```
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (29): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
```

```
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
```

```
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (30-31): 2 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
```

```
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
```

```
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
          )
          (merger): PatchMerger(
            (ln_q): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
            (mlp): Sequential(
              (0): Linear(in_features=5120, out_features=5120,
bias=True)
              (1): GELU(approximate='none')
              (2): Linear(in_features=5120, out_features=3584,
bias=True)
            )
          )
        )
        (language_model): Qwen2VLTextModel(
          (embed_tokens): Embedding(152064, 3584, padding_idx=151654)
          (layers): ModuleList(
            (0-27): 28 x Qwen2VLDecoderLayer(
              (self_attn): Qwen2VLAttention(
                (q_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=3584, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
```

```
                (k_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=512, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=512, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (v_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=512, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=512, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (o_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=3584, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
```

```
                    )
                    (lora_embedding_A): ParameterDict()
                    (lora_embedding_B): ParameterDict()
                    (lora_magnitude_vector): ModuleDict()
                  )
                  (rotary_emb): Qwen2VLRotaryEmbedding()
                )
                (mlp): Qwen2MLP(
                  (gate_proj): lora.Linear4bit(
                    (base_layer): Linear4bit(in_features=3584,
out_features=18944, bias=False)
                    (lora_dropout): ModuleDict(
                      (default): Identity()
                    )
                    (lora_A): ModuleDict(
                      (default): Linear(in_features=3584,
out_features=16, bias=False)
                    )
                    (lora_B): ModuleDict(
                      (default): Linear(in_features=16,
out_features=18944, bias=False)
                    )
                    (lora_embedding_A): ParameterDict()
                    (lora_embedding_B): ParameterDict()
                    (lora_magnitude_vector): ModuleDict()
                  )
                  (up_proj): lora.Linear4bit(
                    (base_layer): Linear4bit(in_features=3584,
out_features=18944, bias=False)
                    (lora_dropout): ModuleDict(
                      (default): Identity()
                    )
                    (lora_A): ModuleDict(
                      (default): Linear(in_features=3584,
out_features=16, bias=False)
                    )
                    (lora_B): ModuleDict(
                      (default): Linear(in_features=16,
out_features=18944, bias=False)
                    )
                    (lora_embedding_A): ParameterDict()
                    (lora_embedding_B): ParameterDict()
                    (lora_magnitude_vector): ModuleDict()
                  )
                  (down_proj): lora.Linear4bit(
                    (base_layer): Linear4bit(in_features=18944,
out_features=3584, bias=False)
                    (lora_dropout): ModuleDict(
                      (default): Identity()
```

```
                    )
                    (lora_A): ModuleDict(
                      (default): Linear(in_features=18944,
out_features=16, bias=False)
                    )
                    (lora_B): ModuleDict(
                      (default): Linear(in_features=16,
out_features=3584, bias=False)
                    )
                    (lora_embedding_A): ParameterDict()
                    (lora_embedding_B): ParameterDict()
                    (lora_magnitude_vector): ModuleDict()
                  )
                  (act_fn): SiLU()
                )
                (input_layernorm): Qwen2RMSNorm((3584,), eps=1e-06)
                (post_attention_layernorm): Qwen2RMSNorm((3584,),
eps=1e-06)
              )
            )
            (norm): Qwen2RMSNorm((3584,), eps=1e-06)
            (rotary_emb): Qwen2VLRotaryEmbedding()
          )
        )
        (lm_head): Linear(in_features=3584, out_features=152064,
bias=False)
      )
    )
)

trainer = SFTTrainer(
    model = model,
    tokenizer=tokenizer,
    data_collator = UnslothVisionDataCollator(model, tokenizer),
    train_dataset = converted_dataset,
    args = SFTConfig(
        per_device_train_batch_size = 2,
        gradient_accumulation_steps = 4,
        warmup_steps = 5,
        max_steps=30,
        learning_rate = 2e-4,
        fp16=not is_bf16_supported(),
        bf16=is_bf16_supported(),
        optim = "adamw_8bit",
        lr_scheduler_type = "linear",
        seed=3407,
        output_dir="outputs",
        report_to = "none",
        remove_unused_columns=False,
        dataset_text_field="",
```

```
        dataset_kwargs = {"skip_prepare_dataset": True},
        dataset_num_proc=4,
        max_seq_length=2048,

    ),
)
```

```
Unsloth: Model does not have a default image size - using 512
```

```
trainer.train()
```

```
The tokenizer has new PAD/BOS/EOS tokens that differ from the model
config and generation config. The model config and generation config
were aligned accordingly, being updated with the tokenizer's values.
Updated tokens: {'bos_token_id': None}.
==((====))==  Unsloth - 2x faster free finetuning | Num GPUs used = 1
   \\   /|    Num examples = 68,686 | Num Epochs = 1 | Total steps =
30
O^O/ \_/ \    Batch size per device = 2 | Gradient accumulation steps
= 4
\        /    Data Parallel GPUs = 1 | Total batch size (2 x 4 x 1) =
8
 "-____-"     Trainable parameters = 50,855,936 of 8,342,231,552
(0.61% trained)
```

```
Unsloth: Will smartly offload gradients to save VRAM!
```

```
<IPython.core.display.HTML object>
```

```
TrainOutput(global_step=30, training_loss=0.4035750709474087,
metrics={'train_runtime': 229.7428, 'train_samples_per_second': 1.045,
'train_steps_per_second': 0.131, 'total_flos': 1684389715636224.0,
'train_loss': 0.4035750709474087, 'epoch': 0.0034941618379291267})
```

```
FastVisionModel.for_inference(model)
```

```
PeftModelForCausalLM(
  (base_model): LoraModel(
    (model): Qwen2VLForConditionalGeneration(
      (model): Qwen2VLModel(
        (visual): Qwen2VisionTransformerPretrainedModel(
          (patch_embed): PatchEmbed(
            (proj): Conv3d(3, 1280, kernel_size=(2, 14, 14),
stride=(2, 14, 14), bias=False)
          )
          (rotary_pos_emb): VisionRotaryEmbedding()
          (blocks): ModuleList(
            (0-18): 19 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
```

```
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
```

```
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (19): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
```

```
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
```

```
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (20-21): 2 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
```

```
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
```

```
            (22): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
```

```
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (23-28): 6 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
```

```
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
```

```
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
            (29): Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
```

```
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=1280,
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
```

```
                )
              )
            )
            (30-31): 2 x Qwen2VLVisionBlock(
              (norm1): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (norm2): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
              (attn): VisionAttention(
                (qkv): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=3840, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3840, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (proj): lora.Linear(
                  (base_layer): Linear(in_features=1280,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
              (mlp): VisionMlp(
                (fc1): lora.Linear(
                  (base_layer): Linear(in_features=1280,
```

```
out_features=5120, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=1280,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=5120, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act): QuickGELUActivation()
                (fc2): lora.Linear(
                  (base_layer): Linear(in_features=5120,
out_features=1280, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=5120,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=1280, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
              )
            )
          )
          (merger): PatchMerger(
            (ln_q): LayerNorm((1280,), eps=1e-06,
elementwise_affine=True)
            (mlp): Sequential(
              (0): Linear(in_features=5120, out_features=5120,
bias=True)
              (1): GELU(approximate='none')
              (2): Linear(in_features=5120, out_features=3584,
bias=True)
            )
          )
```

```
        )
        (language_model): Qwen2VLTextModel(
          (embed_tokens): Embedding(152064, 3584, padding_idx=151654)
          (layers): ModuleList(
            (0-27): 28 x Qwen2VLDecoderLayer(
              (self_attn): Qwen2VLAttention(
                (q_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=3584, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (k_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=512, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=512, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (v_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=512, bias=True)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
```

```
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=512, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (o_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=3584, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (rotary_emb): Qwen2VLRotaryEmbedding()
              )
              (mlp): Qwen2MLP(
                (gate_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=18944, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=18944, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
```

```
                )
                (up_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=3584,
out_features=18944, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=3584,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=18944, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (down_proj): lora.Linear4bit(
                  (base_layer): Linear4bit(in_features=18944,
out_features=3584, bias=False)
                  (lora_dropout): ModuleDict(
                    (default): Identity()
                  )
                  (lora_A): ModuleDict(
                    (default): Linear(in_features=18944,
out_features=16, bias=False)
                  )
                  (lora_B): ModuleDict(
                    (default): Linear(in_features=16,
out_features=3584, bias=False)
                  )
                  (lora_embedding_A): ParameterDict()
                  (lora_embedding_B): ParameterDict()
                  (lora_magnitude_vector): ModuleDict()
                )
                (act_fn): SiLU()
              )
              (input_layernorm): Qwen2RMSNorm((3584,), eps=1e-06)
              (post_attention_layernorm): Qwen2RMSNorm((3584,),
eps=1e-06)
            )
          )
          (norm): Qwen2RMSNorm((3584,), eps=1e-06)
          (rotary_emb): Qwen2VLRotaryEmbedding()
        )
      )
      (lm_head): Linear(in_features=3584, out_features=152064,
bias=False)
```

```
    )
  )
)

image = dataset[2]["image"]

instruction = "Write the Latex representation for this image."

messages = [
    {
        "role": "user",
        "content": [
            {"type": "text", "text": instruction},
            {"type": "image", "image": image}
        ]
    }
]

input_text = tokenizer.apply_chat_template(messages,
add_generation_prompt=True)
input = tokenizer(
    image,
    input_text,
    add_special_tokens = False,
    return_tensors = "pt",

).to("cuda")

from transformers import TextStreamer
text_streamer = TextStreamer(tokenizer, skip_prompt=True)

_ = model.generate(**input, streamer= text_streamer, max_new_tokens =
128, use_cache=True, temperature=1.5, min_p=0.1)

H ^ { \prime } = \beta N \int d \lambda \left\{ \frac { 1 } { 2 \beta
^ { 2 } N ^ { 2 } } \partial _ { \lambda } \zeta ^ { \dagger } \
partial _ { \lambda } \zeta + V ( \lambda ) \zeta ^ { \dagger } \
zeta \right\} \ .<|im_end|>

image
```

$$H' = \beta N \int d\lambda \left\{ \frac{1}{2\beta^2 N^2} \partial_\lambda \zeta^\dagger \partial_\lambda \zeta + V(\lambda) \zeta^\dagger \zeta \right\} .$$