Library Imports

```
! sudo apt-get install -y python3-dev graphviz libgraphviz-dev pkg-config  grap
! pip uninstall -y sentence-transformers
! pip install pygraphviz torch==2.1.0 transformers==4.25.1 torchvision==0.16.0

import torch
import torch.fx
import accelerate
from transformers.utils.fx import (
    symbolic_trace as symbolic_trace_transformers,
)
import transformers
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
warnings.filterwarnings("ignore", category=UserWarning)

# This will reload the imported modules (e.g. analysis) every time you execute
%load_ext autoreload
%autoreload 2
```

```
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
Note, selecting 'libgraphviz-dev' instead of 'graphviz-dev'
pkg-config is already the newest version (0.29.2-1ubuntu3).
python3-dev is already the newest version (3.10.6-1~22.04.1).
graphviz is already the newest version (2.42.2-6ubuntu0.1).
libgraphviz-dev is already the newest version (2.42.2-6ubuntu0.1).
The following packages were automatically installed and are no longer requi
  libbz2-dev libpkgconf3 libreadline-dev
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 29 not upgraded.
WARNING: Skipping sentence-transformers as it is not installed.
Requirement already satisfied: pygraphviz in /usr/local/lib/python3.11/dist
Requirement already satisfied: torch==2.1.0 in /usr/local/lib/python3.11/di
Requirement already satisfied: transformers==4.25.1 in /usr/local/lib/pytho
Requirement already satisfied: torchvision==0.16.0 in /usr/local/lib/python
Requirement already satisfied: torchaudio==2.1.0 in /usr/local/lib/python3.
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.
Requirement already satisfied: sympy in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: networkx in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: jinja2 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: fsspec in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in /usr/loc
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in /usr/l
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in /usr/loc
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in /usr/local/li
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in /usr/local/l
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in /usr/local/l
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in /usr/local
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in /usr/loc
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in /usr/loc
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in /usr/loc
Requirement already satisfied: nvidia-nccl-cu12==2.18.1 in /usr/local/lib/p
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in /usr/local/lib
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.11/d
Requirement already satisfied: huggingface-hub<1.0,>=0.10.0 in /usr/local/l
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dis
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dis
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in /usr/lo
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist
Requirement already satisfied: pillow!=8.3.*,>=5.3.0 in /usr/local/lib/pyth
Requirement already satisfied: nvidia-nvjitlink-cu12 in /usr/local/lib/pyth
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/p
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/di
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3
```

Initialization: Understanding FX Graphs

The **symbolic tracer** in PyTorch FX performs “symbolic execution” of Python code, breaking down a function into individual operations.

How It Works

- The tracer records function execution **symbolically**, capturing all operations into an **FX graph**.
- This enables transformation and analysis of the computational graph before execution.

Example

Below is a dummy module with various operators, followed by its corresponding **FX graph node table**.

```
from analysis import visualize, trace

# Simple module definition
class ModuleA1(torch.nn.Module):
    def __init__(self):
        super().__init__()
        self.param1 = torch.nn.Parameter(torch.rand(30, 120))
        self.param2 = torch.nn.Parameter(torch.rand(30, 120))
        self.linear = torch.nn.Linear(120, 10)

    def forward(self, x):
        x = torch.mul(x, self.param1)
        x = torch.mul(x, self.param2)
        x = self.linear(x)
        return x.relu()
```

```
# Module declaration
module = ModuleA1()

# FX graph.
module_graph = torch.fx.symbolic_trace(module)

# Print the graph as table.
print("Graph Table:")
module_graph.graph.print_tabular()
```

```
Graph Table:
opcode         name    target
-------------  ------  -----------------------------------------------------
placeholder    x       x
get_attr       param1  param1
call_function  mul     <built-in method mul of type object at 0x7bc135e97aa
get_attr       param2  param2
call_function  mul_1   <built-in method mul of type object at 0x7bc135e97aa
call_module    linear  linear
call_method    relu    relu
output         output  output
```

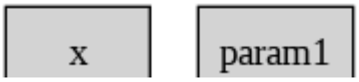
A **Graph** is a data structure representing a `GraphModule` method. It contains the following key elements:

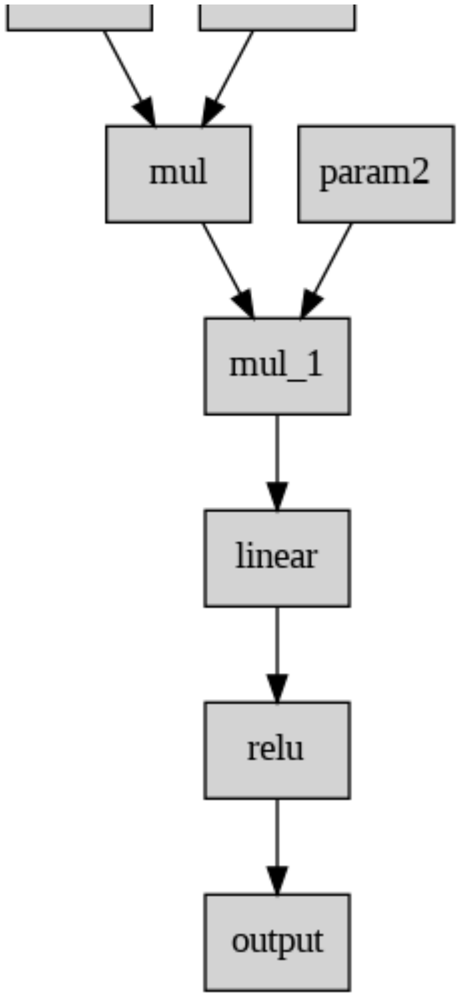
- **Inputs (`placeholder` nodes)**:
  - In FX, method inputs are represented by special `placeholder` nodes.
  - In this case, we have a single `placeholder` node with a target of `x`, meaning the method has one (non-self) argument named `x`.
- **Operation Nodes**:
  - The `get_attr`, `call_function`, `call_module`, and `call_method` nodes represent different operations within the method.
  - A detailed explanation of each can be found in the Node documentation.
- **Output (`output` node)**:
  - The return value in a Graph is represented by a special `output` node.

Refer to the linked documentation (Node, Graph) for a deeper understanding of FX graph structures before proceeding further.

```
# Graph visualization
print("Graph:")
visualize(module_graph)
```

```
Graph:
```





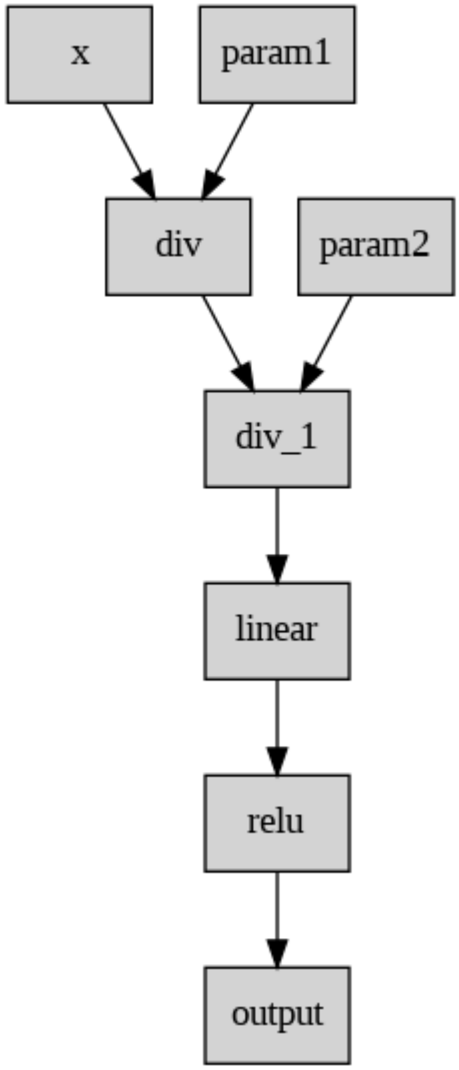
A1: Graph Manipulation [1 point]

You can modify an FX graph by replacing a node's target function. Complete the `transform` function to modify existing graph nodes:

- Replace all nodes using the `torch.mul` operator with `torch.div`.

```
def transform(m: torch.nn.Module,
              tracer_class : type = torch.fx.Tracer) -> torch.nn.Module:
    # FX represents its Graph as an ordered list of nodes so we can iterate thro
    graph : torch.fx.Graph = tracer_class().trace(m)
    for node in graph.nodes:
      if node.target == torch.mul:
        node.target = torch.div

        ## Write the code to replace nodes with mul operator by div operator.
        # pass

    gm = torch.fx.GraphModule(m, graph)
    gm.recompile()
    return gm
```

```
new_model = transform(module)
new_module_graph = torch.fx.symbolic_trace(new_model)
visualize(new_module_graph)
new_module_graph.graph.print_tabular()
```



```
opcode         name    target
-------------  ------  -----------------------------------------------------
placeholder    x       x
get_attr       param1  param1
call_function  div     <built-in method div of type object at 0x7e250be97aa
get_attr       param2  param2
call_function  div_1   <built-in method div of type object at 0x7e250be97aa
call_module    linear  linear
call_method    relu    relu
output         output  output
```

A2: Graph Analysis I [6 points]

Next, we will analyze a trace graph of one layer of the GPT-3 2.7B model.
Review the `analysis.py` file and complete the following:

1. Set `node.shape` in `NodeProp`

- Assign `node.shape` to the output shape of the node.

- Store this value as a **list**, following the format of `torch.Tensor.shape` output.
- If the output is **not** a tensor, set `node.shape = None`.

2. Set `node.latency` in `NodeProp`

- Measure the latency of each operator by running it **10 times** and averaging the results.
- Use `time.time()` to record execution time and convert the value to **milliseconds**.
- For the operators **"placeholder," "output," and "get_attr"**, set `node.latency = 0`.

```
from model import load_model, input_provider
from analysis import NodeProp, visualize, trace, print_graph, dump_graph

model = load_model()  ## loads the GPT model.
input_ids = input_provider(micro_batch_size=1, sequence_length=1024)  ## Dummy

## Graph Generation
graphmodule = trace(model)

print_graph(graphmodule) # Use this for debugging

nodeprop = NodeProp(graphmodule)
graphmodule = nodeprop.propagate(input_ids)

## Graph Visualization
visualize(graphmodule)

# This line creates csv file for submission
dump_graph(graphmodule)
```

```
{'op': 'placeholder', 'target': 'input_ids', 'args': (), 'kwargs': {}}
{'op': 'call_method', 'target': 'size', 'args': (input_ids,), 'kwargs': {}}
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_method', 'target': 'view', 'args': (input_ids, -1, getitem), '
{'op': 'call_method', 'target': 'size', 'args': (view,), 'kwargs': {}}
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (vie
{'op': 'call_function', 'target': <built-in method arange of type object at
{'op': 'call_method', 'target': 'unsqueeze', 'args': (arange, 0), 'kwargs':
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_method', 'target': 'view', 'args': (unsqueeze, -1, getitem_3),
{'op': 'call_module', 'target': 'transformer.wte', 'args': (view,), 'kwargs
{'op': 'call_module', 'target': 'transformer.wpe', 'args': (view_1,), 'kwar
{'op': 'call_function', 'target': <built-in function add>, 'args': (transfo
{'op': 'call_module', 'target': 'transformer.drop', 'args': (add_1,), 'kwar
{'op': 'call_method', 'target': 'size', 'args': (transformer_drop, -1), 'kw
{'op': 'call_function', 'target': <built-in function add>, 'args': (size, (
{'op': 'call_module', 'target': 'transformer.h.0.ln_1', 'args': (transforme
{'op': 'call_method', 'target': 'size', 'args': (transformer_h_0_ln_1,), 'k
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'get_attr', 'target': 'transformer.h.0.attn.c_attn.bias', 'args': ()
```

```
{'op': 'call_method', 'target': 'size', 'args': (transformer_h_0_ln_1, -1),
{'op': 'call_method', 'target': 'view', 'args': (transformer_h_0_ln_1, -1,
{'op': 'get_attr', 'target': 'transformer.h.0.attn.c_attn.weight', 'args':
{'op': 'call_function', 'target': <built-in method addmm of type object at
{'op': 'call_method', 'target': 'view', 'args': (addmm, add_3), 'kwargs': {
{'op': 'call_method', 'target': 'split', 'args': (view_3, 2560), 'kwargs':
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (spl
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (spl
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (spl
{'op': 'call_method', 'target': 'size', 'args': (getitem_5,), 'kwargs': {}}
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'call_method', 'target': 'view', 'args': (getitem_5, add_4), 'kwargs
{'op': 'call_method', 'target': 'permute', 'args': (view_4, 0, 2, 1, 3), 'k
{'op': 'call_method', 'target': 'size', 'args': (getitem_6,), 'kwargs': {}}
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'call_method', 'target': 'view', 'args': (getitem_6, add_5), 'kwargs
{'op': 'call_method', 'target': 'permute', 'args': (view_5, 0, 2, 1, 3), 'k
{'op': 'call_method', 'target': 'size', 'args': (getitem_7,), 'kwargs': {}}
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'call_method', 'target': 'view', 'args': (getitem_7, add_6), 'kwargs
{'op': 'call_method', 'target': 'permute', 'args': (view_6, 0, 2, 1, 3), 'k
{'op': 'call_method', 'target': 'transpose', 'args': (permute_1, -1, -2), '
{'op': 'call_function', 'target': <built-in method matmul of type object at
{'op': 'call_method', 'target': 'size', 'args': (permute_2, -1), 'kwargs':
{'op': 'call_function', 'target': <built-in function pow>, 'args': (size_8,
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (mat
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (mat
{'op': 'call_function', 'target': <built-in method full of type object at 0
{'op': 'call_function', 'target': <built-in function truediv>, 'args': (mat
{'op': 'call_method', 'target': 'size', 'args': (permute, -2), 'kwargs': {}
{'op': 'call_method', 'target': 'size', 'args': (permute_1, -2), 'kwargs':
{'op': 'get_attr', 'target': 'transformer.h.0.attn.bias', 'args': (), 'kwar
{'op': 'call_function', 'target': <built-in function sub>, 'args': (size_10
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (tra
{'op': 'call_method', 'target': 'to', 'args': (getitem_11, torch.bool), 'kw
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (tru
{'op': 'call_function', 'target': <class 'torch.finfo'>, 'args': (getattr_4
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (fin
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (tru
{'op': 'call_function', 'target': <built-in method full of type object at 0
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (tru
{'op': 'call_method', 'target': 'to', 'args': (full_1, getattr_7), 'kwargs'
{'op': 'call_function', 'target': <built-in method where of type object at
{'op': 'call_function', 'target': <function softmax at 0x7bc134166160>, 'ar
{'op': 'call_function', 'target': <built-in function getattr>, 'args': (per
{'op': 'call_method', 'target': 'type', 'args': (softmax, getattr_8), 'kwar
{'op': 'call_module', 'target': 'transformer.h.0.attn.attn_dropout', 'args'
{'op': 'call_function', 'target': <built-in method matmul of type object at
{'op': 'call_method', 'target': 'permute', 'args': (matmul_1, 0, 2, 1, 3),
{'op': 'call_method', 'target': 'contiguous', 'args': (permute_3,), 'kwargs
{'op': 'call_method', 'target': 'size', 'args': (contiguous,), 'kwargs': {}
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'call_method', 'target': 'view', 'args': (contiguous, add_7), 'kwarg
{'op': 'call_method', 'target': 'size', 'args': (view_7,), 'kwargs': {}}
```

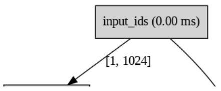
```
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'get_attr', 'target': 'transformer.h.0.attn.c_proj.bias', 'args': ()
{'op': 'call_method', 'target': 'size', 'args': (view_7, -1), 'kwargs': {}}
{'op': 'call_method', 'target': 'view', 'args': (view_7, -1, size_13), 'kwa
{'op': 'get_attr', 'target': 'transformer.h.0.attn.c_proj.weight', 'args':
{'op': 'call_function', 'target': <built-in method addmm of type object at
{'op': 'call_method', 'target': 'view', 'args': (addmm_1, add_8), 'kwargs':
{'op': 'call_module', 'target': 'transformer.h.0.attn.resid_dropout', 'args
{'op': 'call_function', 'target': <built-in function add>, 'args': (transfo
{'op': 'call_module', 'target': 'transformer.h.0.ln_2', 'args': (add_9,), '
{'op': 'call_method', 'target': 'size', 'args': (transformer_h_0_ln_2,), 'k
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'get_attr', 'target': 'transformer.h.0.mlp.c_fc.bias', 'args': (), '
{'op': 'call_method', 'target': 'size', 'args': (transformer_h_0_ln_2, -1),
{'op': 'call_method', 'target': 'view', 'args': (transformer_h_0_ln_2, -1,
{'op': 'get_attr', 'target': 'transformer.h.0.mlp.c_fc.weight', 'args': (),
{'op': 'call_function', 'target': <built-in method addmm of type object at
{'op': 'call_method', 'target': 'view', 'args': (addmm_2, add_10), 'kwargs'
{'op': 'call_function', 'target': <built-in function mul>, 'args': (0.5, vi
{'op': 'call_function', 'target': <built-in method pow of type object at 0x
{'op': 'call_function', 'target': <built-in function mul>, 'args': (0.04471
{'op': 'call_function', 'target': <built-in function add>, 'args': (view_11
{'op': 'call_function', 'target': <built-in function mul>, 'args': (0.79788
{'op': 'call_function', 'target': <built-in method tanh of type object at 0
{'op': 'call_function', 'target': <built-in function add>, 'args': (1.0, ta
{'op': 'call_function', 'target': <built-in function mul>, 'args': (mul, ad
{'op': 'call_method', 'target': 'size', 'args': (mul_3,), 'kwargs': {}}
{'op': 'call_function', 'target': <built-in function getitem>, 'args': (siz
{'op': 'call_function', 'target': <built-in function add>, 'args': (getitem
{'op': 'get_attr', 'target': 'transformer.h.0.mlp.c_proj.bias', 'args': (),
{'op': 'call_method', 'target': 'size', 'args': (mul_3, -1), 'kwargs': {}}
{'op': 'call_method', 'target': 'view', 'args': (mul_3, -1, size_17), 'kwar
{'op': 'get_attr', 'target': 'transformer.h.0.mlp.c_proj.weight', 'args': (
{'op': 'call_function', 'target': <built-in method addmm of type object at
{'op': 'call_method', 'target': 'view', 'args': (addmm_3, add_13), 'kwargs'
{'op': 'call_module', 'target': 'transformer.h.0.mlp.dropout', 'args': (vie
{'op': 'call_function', 'target': <built-in function add>, 'args': (add_9,
{'op': 'call_module', 'target': 'transformer.ln_f', 'args': (add_14,), 'kwa
{'op': 'call_method', 'target': 'view', 'args': (transformer_ln_f, add_2),
{'op': 'call_module', 'target': 'lm_head', 'args': (view_14,), 'kwargs': {}
{'op': 'output', 'target': 'output', 'args': ({'logits': lm_head, 'past_key
{input_ids: tensor([[1, 1, 1,  ..., 1, 1, 1]], dtype=torch.int32), size: to
         [ 0.0121, -0.0233, -0.0008,  ...,  0.0108, -0.0097, -0.0323],
         [ 0.0121, -0.0233, -0.0008,  ...,  0.0108, -0.0097, -0.0323],
         ...,
         [ 0.0121, -0.0233, -0.0008,  ...,  0.0108, -0.0097, -0.0323],
         [ 0.0121, -0.0233, -0.0008,  ...,  0.0108, -0.0097, -0.0323],
         [ 0.0121, -0.0233, -0.0008,  ...,  0.0108, -0.0097, -0.0323]]],
       grad_fn=<EmbeddingBackward0>), transformer_wpe: tensor([[[ 0.0170,
         [-0.0141, -0.0033, -0.0224,  ...,  0.0064,  0.0235,  0.0164],
         [ 0.0049, -0.0300,  0.0265,  ..., -0.0188,  0.0284, -0.0369],
         ...,
         [-0.0127, -0.0152,  0.0364,  ..., -0.0177,  0.0203, -0.0173],
         [-0.0093, -0.0035, -0.0343,  ..., -0.0298, -0.0015,  0.0258],
         [ 0.0047, -0.0314, -0.0080,  ...,  0.0316,  0.0050,  0.0028]]],
       grad_fn=<EmbeddingBackward0>), add_1: tensor([[[ 0.0291,  0.0047,  0
```

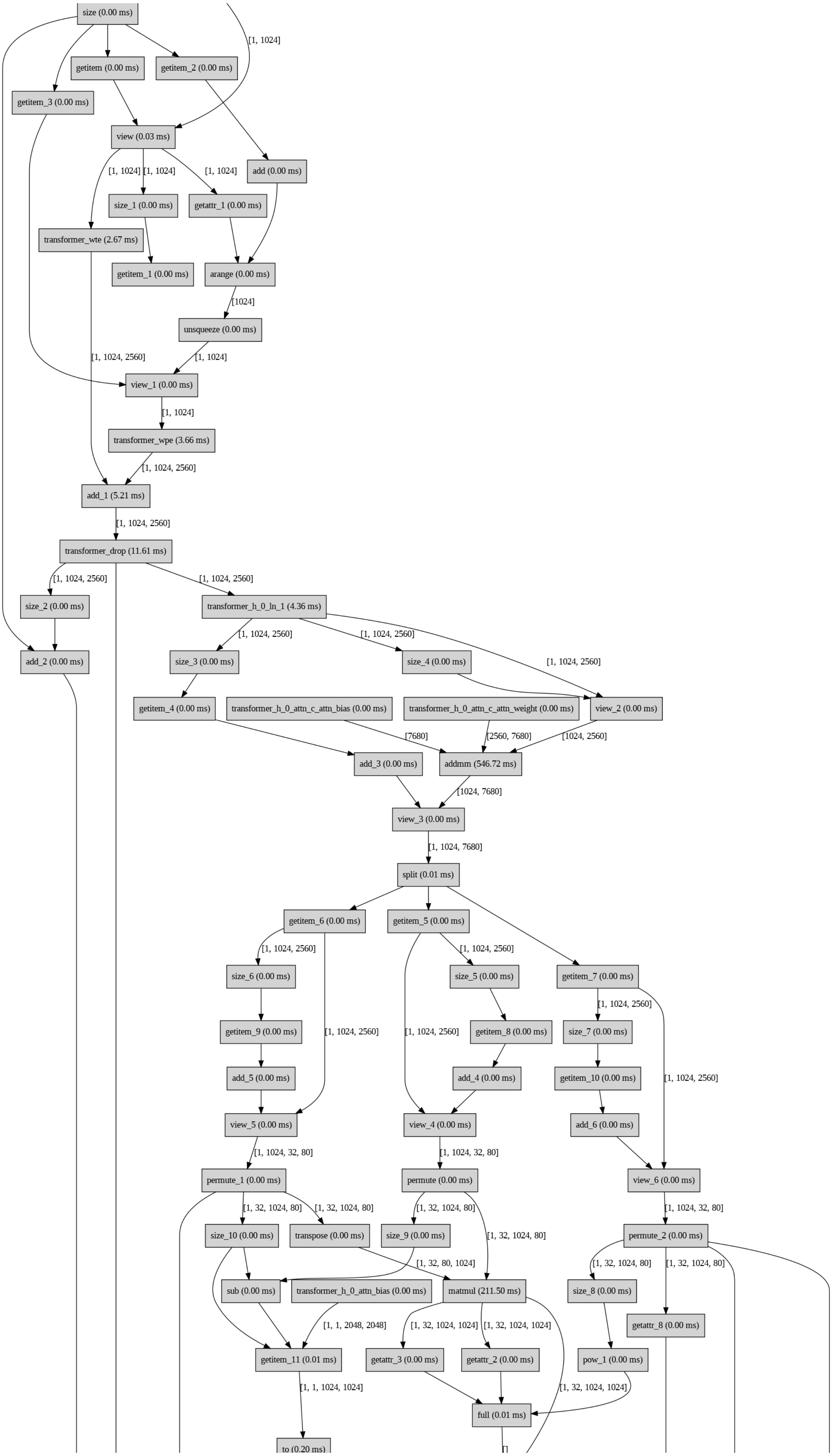
```
  [-0.0021, -0.0266, -0.0232,  ...,  0.0172,  0.0138, -0.0159],
  [ 0.0170, -0.0534,  0.0257,  ..., -0.0080,  0.0187, -0.0692],
  ...,
  [-0.0006, -0.0386,  0.0356,  ..., -0.0070,  0.0106, -0.0495],
  [ 0.0028, -0.0268, -0.0352,  ..., -0.0190, -0.0112, -0.0065],
  [ 0.0168, -0.0547, -0.0088,  ...,  0.0424, -0.0047, -0.0294]]],
grad_fn=<AddBackward0>), transformer_drop: tensor([[[ 0.0323,  0.005
  [-0.0000, -0.0296, -0.0258,  ...,  0.0191,  0.0153, -0.0177],
  [ 0.0189, -0.0593,  0.0285,  ..., -0.0089,  0.0207, -0.0000],
  ...,
  [-0.0006, -0.0429,  0.0396,  ..., -0.0077,  0.0118, -0.0550],
  [ 0.0031, -0.0298, -0.0391,  ..., -0.0211, -0.0124, -0.0073],
  [ 0.0187, -0.0608, -0.0098,  ...,  0.0471, -0.0052, -0.0327]]],
grad_fn=<MulBackward0>), size_2: 2560, add_2: torch.Size([1, 1024, 2
  [-0.0139, -0.9942, -0.8683,  ...,  0.6186,  0.4942, -0.6006],
  [ 0.6456, -1.9920,  0.9696,  ..., -0.2918,  0.7072,  0.0074],
  ...,
  [-0.0208, -1.4737,  1.3632,  ..., -0.2646,  0.4081, -1.8936],
  [ 0.0806, -1.0072, -1.3144,  ..., -0.7194, -0.4321, -0.2624],
  [ 0.6092, -2.0067, -0.3285,  ...,  1.5452, -0.1758, -1.0820]]],
grad_fn=<NativeLayerNormBackward0>), size_3: torch.Size([1, 1024, 25
 [-0.0139, -0.9942, -0.8683,  ...,  0.6186,  0.4942, -0.6006],
 [ 0.6456, -1.9920,  0.9696,  ..., -0.2918,  0.7072,  0.0074],
 ...,
 [-0.0208, -1.4737,  1.3632,  ..., -0.2646,  0.4081, -1.8936],
 [ 0.0806, -1.0072, -1.3144,  ..., -0.7194, -0.4321, -0.2624],
 [ 0.6092, -2.0067, -0.3285,  ...,  1.5452, -0.1758, -1.0820]],
grad_fn=<ViewBackward0>), transformer_h_0_attn_c_attn_weight: tensor
 [-0.0137,  0.0036, -0.0325,  ..., -0.0308, -0.0027, -0.0017],
 [ 0.0032,  0.0298,  0.0324,  ...,  0.0096, -0.0262, -0.0263],
 ...,
 [-0.0176, -0.0332,  0.0239,  ...,  0.0316, -0.0208,  0.0212],
 [-0.0079, -0.0137, -0.0593,  ...,  0.0019, -0.0119,  0.0157],
 [-0.0198,  0.0130, -0.0247,  ...,  0.0070, -0.0028,  0.0066]]), add
   1.2549e-01,  9.6314e-01],
 [ 4.4119e-01,  9.4756e-01,  1.9806e+00,  ..., -1.5521e-04,
  -1.2065e-01,  3.4900e-01],
 [-6.5400e-01,  1.7180e+00,  1.5663e+00,  ...,  3.2143e-01,
   1.0082e+00, -6.9631e-01],
 ...,
 [-5.1544e-01,  1.2657e+00,  8.5728e-01,  ..., -5.6343e-01,
  -8.9841e-01,  8.3196e-01],
 [-5.8501e-01,  1.0318e+00,  2.1008e+00,  ...,  5.7716e-01,
   7.4068e-01,  1.7688e-01],
 [ 1.5093e-01,  1.4387e+00,  9.6359e-01,  ...,  7.7132e-01,
   2.3493e-01,  1.0911e+00]], grad_fn=<AddmmBackward0>), view_3: ten
    1.2549e-01,  9.6314e-01],
  [ 4.4119e-01,  9.4756e-01,  1.9806e+00,  ..., -1.5521e-04,
   -1.2065e-01,  3.4900e-01],
  [-6.5400e-01,  1.7180e+00,  1.5663e+00,  ...,  3.2143e-01,
    1.0082e+00, -6.9631e-01],
  ...,
  [-5.1544e-01,  1.2657e+00,  8.5728e-01,  ..., -5.6343e-01,
   -8.9841e-01,  8.3196e-01],
  [-5.8501e-01,  1.0318e+00,  2.1008e+00,  ...,  5.7716e-01,
    7.4068e-01,  1.7688e-01],
  [ 1.5093e-01,  1.4387e+00,  9.6359e-01,  ...,  7.7132e-01,
    2.3493e-01,  1.0911e+00]]], grad_fn=<ViewBackward0>), split: (te
  [ 0.4412,  0.9476,  1.9806,  ...,  0.2939, -0.9595, -0.5674],
  [-0.6540,  1.7180,  1.5663,  ...,  0.8883,  0.0366, -1.4258],
  ...,
  [-0.5154,  1.2657,  0.8573,  ..., -0.8223, -0.5740, -1.5449],
  [-0.5850,  1.0318,  2.1008,  ..., -1.2132, -1.1229,  0.5343],
  [ 0.1509,  1.4387,  0.9636,  ...,  1.1912, -0.7553, -1.8944]]],
grad_fn=<SplitBackward0>), tensor([[[-0.0947,  0.8811, -0.1241,  ...
  [-0.2898, -0.9153,  0.2059,  ...,  1.5720, -0.0148,  0.3059],
  [ 0.8982,  0.5432,  0.0756,  ...,  0.8550,  0.7815, -0.1665],
  ...,
  [ 0.7235, -0.1132, -0.0178,  ...,  0.5835,  0.7090, -1.3823],
  [ 0.4877,  0.9024,  0.1619,  ...,  2.2055,  2.1243,  1.1146],
  [-0.3400, -0.1774,  1.1426,  ...,  1.5862,  1.8645, -0.2542]]],
grad_fn=<SplitBackward0>), tensor([[[ 5.4475e-01,  4.1136e-01,  6.79
    1.2549e-01,  9.6314e-01],
  [ 1.4472e-01,  1.5402e-01, -4.0394e-01,  ..., -1.5521e-04,
   -1.2065e-01,  3.4900e-01],
  [ 1.3474e+00, -1.0317e-02, -6.0596e-01,  ...,  3.2143e-01,
    1.0082e+00, -6.9631e-01],
  ...,
  [ 3.4586e-01, -1.0730e+00,  9.3181e-01,  ..., -5.6343e-01,
   -8.9841e-01,  8.3196e-01],
  [-1.0901e+00, -1.2123e+00, -5.5091e-01,  ...,  5.7716e-01,
    7.4068e-01,  1.7688e-01],
  [ 1.0175e+00, -3.5308e-01,  7.2218e-02,  ...,  7.7132e-01,
    2.3493e-01,  1.0911e+00]]], grad_fn=<SplitBackward0>)), getitem_
  [ 0.4412,  0.9476,  1.9806,  ...,  0.2939, -0.9595, -0.5674],
  [-0.6540,  1.7180,  1.5663,  ...,  0.8883,  0.0366, -1.4258],
  ...,
  [-0.5154,  1.2657,  0.8573,  ..., -0.8223, -0.5740, -1.5449],
  [-0.5850,  1.0318,  2.1008,  ..., -1.2132, -1.1229,  0.5343],
  [ 0.1509,  1.4387,  0.9636,  ...,  1.1912, -0.7553, -1.8944]]],
grad_fn=<SplitBackward0>), getitem_6: tensor([[[-0.0947,  0.8811, -0
  [-0.2898, -0.9153,  0.2059,  ...,  1.5720, -0.0148,  0.3059],
  [ 0.8982,  0.5432,  0.0756,  ...,  0.8550,  0.7815, -0.1665],
  ...,
  [ 0.7235, -0.1132, -0.0178,  ...,  0.5835,  0.7090, -1.3823],
  [ 0.4877,  0.9024,  0.1619,  ...,  2.2055,  2.1243,  1.1146],
  [-0.3400, -0.1774,  1.1426,  ...,  1.5862,  1.8645, -0.2542]]],
grad_fn=<SplitBackward0>), getitem_7: tensor([[[ 5.4475e-01,  4.1136
    1.2549e-01,  9.6314e-01],
  [ 1.4472e-01,  1.5402e-01, -4.0394e-01,  ..., -1.5521e-04,
   -1.2065e-01,  3.4900e-01],
  [ 1.3474e+00, -1.0317e-02, -6.0596e-01,  ...,  3.2143e-01,
    1.0082e+00, -6.9631e-01],
  ...,
  [ 3.4586e-01, -1.0730e+00,  9.3181e-01,  ..., -5.6343e-01,
   -8.9841e-01,  8.3196e-01],
  [-1.0901e+00, -1.2123e+00, -5.5091e-01,  ...,  5.7716e-01,
    7.4068e-01,  1.7688e-01],
  [ 1.0175e+00, -3.5308e-01,  7.2218e-02,  ...,  7.7132e-01,
    2.3493e-01,  1.0911e+00]]], grad_fn=<SplitBackward0>), size_5: t
   -1.2857e+00, -3.1551e-01],
   [-3.2800e-02, -5.6794e-01, -1.8748e-01,  ..., -1.2788e-02,
     1.0524e+00, -9.5777e-01],
   [-1.0945e+00,  7.3084e-01, -1.2461e+00,  ..., -4.8138e-01,
    -1.2116e+00,  1.4660e-01],
   ...,
   [-1.6830e+00,  4.6264e-02, -5.4264e-01,  ...,  1.2484e+00,
    -7.6872e-01, -5.5794e-01],
   [ 5.3973e-01, -1.3272e+00,  1.8654e+00,  ..., -1.2828e+00,
     9.2584e-01,  3.3336e-01],
   [ 1.2195e+00,  4.3383e-01, -1.2964e+00,  ...,  4.8461e-01,
     8.0541e-01,  8.0558e-01]],

  [[ 4.4119e-01,  9.4756e-01,  1.9806e+00,  ...,  5.5374e-01,
     5.3315e-01,  9.1289e-01],
   [-5.5214e-01, -7.9581e-01, -1.8301e-01,  ...,  7.4550e-01,
     2.6868e-01, -1.7955e+00],
   [-5.0503e-02,  1.1604e+00, -6.8851e-01,  ..., -3.1103e-02,
     1.5962e+00,  7.0856e-01],
   ...,
   [-1.4985e+00, -2.7073e-01, -5.0553e-02,  ...,  1.9765e+00,
    -2.1393e+00, -1.8217e+00],
   [-2.1767e-01, -6.2865e-01,  4.9939e-01,  ..., -2.0071e+00,
     1.2923e+00,  1.2966e-01],
   [ 4.6457e-01, -1.5038e-02, -1.1027e+00,  ...,  2.9391e-01,
    -9.5952e-01, -5.6737e-01]],

  [[-6.5400e-01,  1.7180e+00,  1.5663e+00,  ...,  4.5824e-01,
     6.0123e-01,  1.0662e-01],
   [-8.7456e-01, -1.0741e+00,  2.2769e-01,  ..., -1.0926e+00,
     3.6365e-01,  2.8203e-01],
   [-1.2875e+00,  7.2619e-01, -1.2944e+00,  ...,  1.1848e+00,
    -6.4400e-01,  1.1598e+00],
   ...,
   [-7.0545e-01,  6.2366e-02, -3.0401e-01,  ...,  1.4635e+00,
    -1.5522e+00, -1.1085e+00],
   [ 2.1295e-01, -1.4179e+00,  8.7977e-02,  ..., -6.0971e-01,
     5.3258e-01,  2.2649e-02],
   [-2.2376e-01, -8.4801e-01, -7.1912e-02,  ...,  8.8825e-01,
     3.6628e-02, -1.4258e+00]],


  ...,


  [[-5.1544e-01,  1.2657e+00,  8.5728e-01,  ...,  1.5964e+00,
    -4.3527e-01,  1.4352e+00],
   [-8.8676e-02, -6.7323e-01, -1.3165e+00,  ..., -6.5832e-01,
     1.8433e+00, -6.7086e-01],
   [-7.8150e-01,  6.4667e-01, -2.0857e-01,  ...,  5.7460e-01,
    -1.0659e+00, -2.1837e-02],
   ...,
   [-5.0737e-01,  2.5884e-01, -1.6555e-01,  ..., -2.7461e-03,
    -8.9480e-01,  1.3012e-02],
   [ 1.0539e+00, -4.9611e-01, -8.2595e-01,  ...,  3.6097e-01,
    -2.0568e-01, -4.2166e-01],
   [ 2.1486e+00,  1.0775e-01, -3.3563e-01,  ..., -8.2233e-01,
    -5.7397e-01, -1.5449e+00]],

  [[-5.8501e-01,  1.0318e+00,  2.1008e+00,  ...,  4.2044e-01,
    -1.0137e+00,  5.6344e-01],
   [-1.4472e+00, -7.3342e-01, -1.9121e-01,  ..., -2.8799e-01,
    -1.1699e-01, -7.5985e-01],
   [-4.2458e-01, -8.1531e-02,  5.6539e-01,  ..., -1.4848e+00,
    -2.7788e-01, -2.5043e-01],
   ...,
   [ 6.1142e-01,  7.5670e-01,  5.6700e-01,  ...,  1.3557e+00,
   [-6.1143e-01,  7.5670e-01, -5.6799e-01,  ...,  1.3557e+00,
     5.7326e-01, -5.1458e-01],
   [ 6.1721e-01,  3.3020e-01,  1.5350e-01,  ..., -1.8831e-01,
     1.5387e+00, -3.9524e-01],
   [ 6.7597e-01, -4.9223e-01, -7.5036e-01,  ..., -1.2132e+00,
    -1.1229e+00,  5.3432e-01]],

  [[ 1.5093e-01,  1.4387e+00,  9.6359e-01,  ...,  1.5343e+00,
    -4.6784e-01, -6.5154e-01],
   [-1.2458e+00, -1.7348e+00, -5.8817e-01,  ..., -4.3168e-01,
     6.7754e-01, -1.3545e+00],
   [-6.6876e-01, -2.1163e-01,  1.8188e-03,  ...,  1.6177e-01,
     1.7669e-01, -2.0214e-01],
   ...,
   [-1.6581e+00,  1.5965e-01,  7.2994e-01,  ...,  8.9593e-01,
    -9.1164e-01, -8.7880e-01],
   [ 1.3541e-01, -3.0570e-01,  1.2660e+00,  ..., -2.0911e+00,
     2.9748e-01, -6.0567e-01],
   [-3.4520e-01,  3.8611e-01, -1.2274e+00,  ...,  1.1912e+00,
    -7.5532e-01, -1.8944e+00]]]], grad_fn=<ViewBackward0>), permute:
    -1.2857e+00, -3.1551e-01],
   [ 4.4119e-01,  9.4756e-01,  1.9806e+00,  ...,  5.5374e-01,
     5.3315e-01,  9.1289e-01],
   [-6.5400e-01,  1.7180e+00,  1.5663e+00,  ...,  4.5824e-01,
     6.0123e-01,  1.0662e-01],
   ...,
   [-5.1544e-01,  1.2657e+00,  8.5728e-01,  ...,  1.5964e+00,
    -4.3527e-01,  1.4352e+00],
   [-5.8501e-01,  1.0318e+00,  2.1008e+00,  ...,  4.2044e-01,
    -1.0137e+00,  5.6344e-01],
   [ 1.5093e-01,  1.4387e+00,  9.6359e-01,  ...,  1.5343e+00,
    -4.6784e-01, -6.5154e-01]],

  [[-3.2800e-02, -5.6794e-01, -1.8748e-01,  ..., -1.2788e-02,
     1.0524e+00, -9.5777e-01],
   [-5.5214e-01, -7.9581e-01, -1.8301e-01,  ...,  7.4550e-01,
     2.6868e-01, -1.7955e+00],
   [-8.7456e-01, -1.0741e+00,  2.2769e-01,  ..., -1.0926e+00,
     3.6365e-01,  2.8203e-01],
   ...,
   [-8.8676e-02, -6.7323e-01, -1.3165e+00,  ..., -6.5832e-01,
     1.8433e+00, -6.7086e-01],
   [-1.4472e+00, -7.3342e-01, -1.9121e-01,  ..., -2.8799e-01,
    -1.1699e-01, -7.5985e-01],
   [-1.2458e+00, -1.7348e+00, -5.8817e-01,  ..., -4.3168e-01,
     6.7754e-01, -1.3545e+00]],

  [[-1.0945e+00,  7.3084e-01, -1.2461e+00,  ..., -4.8138e-01,
    -1.2116e+00,  1.4660e-01],
   [-5.0503e-02,  1.1604e+00, -6.8851e-01,  ..., -3.1103e-02,
     1.5962e+00,  7.0856e-01],
   [-1.2875e+00,  7.2619e-01, -1.2944e+00,  ...,  1.1848e+00,
    -6.4400e-01,  1.1598e+00],
   ...,
   [-7.8150e-01,  6.4667e-01, -2.0857e-01,  ...,  5.7460e-01,
    -1.0659e+00, -2.1837e-02],
   [-4.2458e-01, -8.1531e-02,  5.6539e-01,  ..., -1.4848e+00,
    -2.7788e-01, -2.5043e-01],
   [-6.6876e-01, -2.1163e-01,  1.8188e-03,  ...,  1.6177e-01,
   [-6.6876e-01, -2.1163e-01,  1.8188e-03,  ...,  1.6177e-01,
     1.7669e-01, -2.0214e-01]],

  ...,

  [[-1.6830e+00,  4.6264e-02, -5.4264e-01,  ...,  1.2484e+00,
    -7.6872e-01, -5.5794e-01],
   [-1.4985e+00, -2.7073e-01, -5.0553e-02,  ...,  1.9765e+00,
    -2.1393e+00, -1.8217e+00],
   [-7.0545e-01,  6.2366e-02, -3.0401e-01,  ...,  1.4635e+00,
    -1.5522e+00, -1.1085e+00],
   ...,
   [-5.0737e-01,  2.5884e-01, -1.6555e-01,  ..., -2.7461e-03,
    -8.9480e-01,  1.3012e-02],
   [-6.1143e-01,  7.5670e-01, -5.6799e-01,  ...,  1.3557e+00,
     5.7326e-01, -5.1458e-01],
   [-1.6581e+00,  1.5965e-01,  7.2994e-01,  ...,  8.9593e-01,
    -9.1164e-01, -8.7880e-01]],

  [[ 5.3973e-01, -1.3272e+00,  1.8654e+00,  ..., -1.2828e+00,
     9.2584e-01,  3.3336e-01],
   [-2.1767e-01, -6.2865e-01,  4.9939e-01,  ..., -2.0071e+00,
     1.2923e+00,  1.2966e-01],
   [ 2.1295e-01, -1.4179e+00,  8.7977e-02,  ..., -6.0971e-01,
     5.3258e-01,  2.2649e-02],
   ...,
   [ 1.0539e+00, -4.9611e-01, -8.2595e-01,  ...,  3.6097e-01,
    -2.0568e-01, -4.2166e-01],
   [ 6.1721e-01,  3.3020e-01,  1.5350e-01,  ..., -1.8831e-01,
     1.5387e+00, -3.9524e-01],
   [ 1.3541e-01, -3.0570e-01,  1.2660e+00,  ..., -2.0911e+00,
     2.9748e-01, -6.0567e-01]],

  [[ 1.2195e+00,  4.3383e-01, -1.2964e+00,  ...,  4.8461e-01,
     8.0541e-01,  8.0558e-01],
   [ 4.6457e-01, -1.5038e-02, -1.1027e+00,  ...,  2.9391e-01,
    -9.5952e-01, -5.6737e-01],
   [-2.2376e-01, -8.4801e-01, -7.1912e-02,  ...,  8.8825e-01,
     3.6628e-02, -1.4258e+00],
   ...,
   [ 2.1486e+00,  1.0775e-01, -3.3563e-01,  ..., -8.2233e-01,
    -5.7397e-01, -1.5449e+00],
   [ 6.7597e-01, -4.9223e-01, -7.5036e-01,  ..., -1.2132e+00,
    -1.1229e+00,  5.3432e-01],
   [-3.4520e-01,  3.8611e-01, -1.2274e+00,  ...,  1.1912e+00,
    -7.5532e-01, -1.8944e+00]]]], grad_fn=<PermuteBackward0>), size_
   [-1.1550, -0.6857, -0.0548,  ...,  0.7911, -1.2452,  1.1871],
   [ 0.9380, -0.1100, -1.4938,  ...,  0.7169,  0.9322,  0.0979],
   ...,
   [ 1.5353, -0.6293,  0.1201,  ...,  1.1989,  0.0139,  0.5774],
   [-1.7977, -0.0972,  0.2891,  ..., -0.4595, -1.0992,  0.6587],
   [ 0.8641,  0.5925,  0.0250,  ...,  0.9172,  0.8187,  0.0937]],

  [[-0.2898, -0.9153,  0.2059,  ..., -0.7287, -1.3031,  0.2322],
   [-1.4128,  0.3902,  0.5208,  ...,  1.5560, -1.4400,  0.8671],
   [-0.2036,  0.7863,  0.6648,  ...,  0.5309, -0.1505,  0.1302],
   ...,
   [-0.1049, -1.6774,  0.6332,  ...,  1.0535, -0.5428, -0.9433],
   [-0.7844,  0.0462, -1.1205,  ...,  0.4437, -0.6814,  1.1193],
   [ 1.1099,  1.4441,  0.7351,  ...,  1.5720, -0.0148,  0.3059]],

  [[ 0.8982,  0.5432,  0.0756,  ...,  1.0897, -1.7253,  0.2846],
   [-0.8195, -1.1852,  0.7765,  ...,  1.4433, -0.8974,  1.2275],
   [ 0.5280, -1.0059,  1.2404,  ...,  2.1611,  0.4663, -2.0759],
   ...,
   [ 1.0090, -0.6773,  0.5736,  ...,  1.5946,  0.2422, -0.2608],
   [-0.6200,  0.4880, -1.5203,  ...,  0.4020, -0.7304, -0.3302],
   [ 0.5323,  0.8540,  1.2001,  ...,  0.8550,  0.7815, -0.1665]],

  ...,

  [[ 0.7235, -0.1132, -0.0178,  ..., -0.8767,  0.9113,  1.0387],
   [-1.3198, -0.0586,  0.6382,  ...,  0.1729, -0.2085,  0.6523],
   [ 0.3867,  1.8113, -0.2210,  ..., -0.1997,  0.8334, -2.3713],
   ...,
   [ 0.9555, -0.3896,  1.3091,  ...,  1.7802,  0.3060,  0.6750],
   [-1.2726, -0.3915,  0.6958,  ...,  0.4698, -1.2266, -0.3855],
   [ 0.1825, -0.2630, -0.8271,  ...,  0.5835,  0.7090, -1.3823]],

  [[ 0.4877,  0.9024,  0.1619,  ..., -0.7393, -1.1445, -0.2149],
   [-0.1504, -0.2212, -0.6004,  ...,  0.8414,  1.0264,  0.3428],
   [-0.2839,  1.5350, -0.5067,  ...,  0.4349,  0.4209,  0.6408],
   ...,
   [ 1.4842, -0.3871,  1.0065,  ...,  2.0689, -1.3078, -0.7658],
   [-0.9818,  0.4404, -0.4635,  ...,  0.2169, -1.1535,  0.0350],
   [-0.5885, -1.2048,  0.7633,  ...,  2.2055,  2.1243,  1.1146]],

  [[-0.3400, -0.1774,  1.1426,  ..., -1.6697,  0.7312,  0.2506],
   [ 0.4139,  0.7714, -0.0446,  ...,  1.4960, -1.7593,  0.9587],
   [-0.2052, -1.3002, -0.2397,  ...,  0.9969,  1.4980, -1.2450],
   ...,
   [-0.1303, -2.1165,  1.7418,  ...,  1.6911,  0.4163, -1.4566],
   [-0.8389, -0.8111, -0.7810,  ..., -0.4379, -0.9150,  1.7493],
   [ 1.2332, -0.3148,  1.6390,  ...,  1.5862,  1.8645, -0.2542]]]],
grad_fn=<ViewBackward0>), permute_1: tensor([[[[-0.0947,  0.8811, -0
   [-0.2898, -0.9153,  0.2059,  ..., -0.7287, -1.3031,  0.2322],
   [ 0.8982,  0.5432,  0.0756,  ...,  1.0897, -1.7253,  0.2846],
   ...,
   [ 0.7235, -0.1132, -0.0178,  ..., -0.8767,  0.9113,  1.0387],
   [ 0.4877,  0.9024,  0.1619,  ..., -0.7393, -1.1445, -0.2149],
   [-0.3400, -0.1774,  1.1426,  ..., -1.6697,  0.7312,  0.2506]],

  [[-1.1550, -0.6857, -0.0548,  ...,  0.7911, -1.2452,  1.1871],
   [-1.4128,  0.3902,  0.5208,  ...,  1.5560, -1.4400,  0.8671],
   [-0.8195, -1.1852,  0.7765,  ...,  1.4433, -0.8974,  1.2275],
   ...,
   [-1.3198, -0.0586,  0.6382,  ...,  0.1729, -0.2085,  0.6523],
   [-0.1504, -0.2212, -0.6004,  ...,  0.8414,  1.0264,  0.3428],
   [ 0.4139,  0.7714, -0.0446,  ...,  1.4960, -1.7593,  0.9587]],

  [[ 0.9380, -0.1100, -1.4938,  ...,  0.7169,  0.9322,  0.0979],
   [-0.2036,  0.7863,  0.6648,  ...,  0.5309, -0.1505,  0.1302],
   [ 0.5280, -1.0059,  1.2404,  ...,  2.1611,  0.4663, -2.0759],
   ...,
   [ 0.3867,  1.8113, -0.2210,  ..., -0.1997,  0.8334, -2.3713],
   [-0.2839,  1.5350, -0.5067,  ...,  0.4349,  0.4209,  0.6408],
   [-0.2052, -1.3002, -0.2397,  ...,  0.9969,  1.4980, -1.2450]],
   [-0.2052, -1.3002, -0.2397,  ...,  0.9969,  1.4980, -1.2450]],

  ...,

  [[ 1.5353, -0.6293,  0.1201,  ...,  1.1989,  0.0139,  0.5774],
   [-0.1049, -1.6774,  0.6332,  ...,  1.0535, -0.5428, -0.9433],
   [ 1.0090, -0.6773,  0.5736,  ...,  1.5946,  0.2422, -0.2608],
   ...,
   [ 0.9555, -0.3896,  1.3091,  ...,  1.7802,  0.3060,  0.6750],
   [ 1.4842, -0.3871,  1.0065,  ...,  2.0689, -1.3078, -0.7658],
   [-0.1303, -2.1165,  1.7418,  ...,  1.6911,  0.4163, -1.4566]],

  [[-1.7977, -0.0972,  0.2891,  ..., -0.4595, -1.0992,  0.6587],
   [-0.7844,  0.0462, -1.1205,  ...,  0.4437, -0.6814,  1.1193],
   [-0.6200,  0.4880, -1.5203,  ...,  0.4020, -0.7304, -0.3302],
   ...,
   [-1.2726, -0.3915,  0.6958,  ...,  0.4698, -1.2266, -0.3855],
   [-0.9818,  0.4404, -0.4635,  ...,  0.2169, -1.1535,  0.0350],
   [-0.8389, -0.8111, -0.7810,  ..., -0.4379, -0.9150,  1.7493]],

  [[ 0.8641,  0.5925,  0.0250,  ...,  0.9172,  0.8187,  0.0937],
   [ 1.1099,  1.4441,  0.7351,  ...,  1.5720, -0.0148,  0.3059],
   [ 0.5323,  0.8540,  1.2001,  ...,  0.8550,  0.7815, -0.1665],
   ...,
   [ 0.1825, -0.2630, -0.8271,  ...,  0.5835,  0.7090, -1.3823],
   [-0.5885, -1.2048,  0.7633,  ...,  2.2055,  2.1243,  1.1146],
   [ 1.2332, -0.3148,  1.6390,  ...,  1.5862,  1.8645, -0.2542]]]],
grad_fn=<PermuteBackward0>), size_7: torch.Size([1, 1024, 2560]), ge
     5.3991e-02, -1.1233e+00],
   [-3.1268e-01,  7.5064e-01,  1.3438e+00,  ..., -1.2461e-01,
     4.1225e-01,  1.0110e+00],
   [ 5.3532e-01, -1.7978e-01, -9.8737e-01,  ...,  7.4715e-01,
    -4.8342e-01,  1.5643e+00],
   ...,
   [-6.7070e-01, -9.6281e-01,  3.8247e-01,  ...,  1.4481e+00,
    -6.7513e-01,  7.8456e-01],
   [ 3.2699e-02, -1.7002e+00,  9.3860e-01,  ...,  1.0049e+00,
     1.3244e+00, -1.6203e+00],
   [-2.1535e-01, -2.6749e+00,  2.0272e-01,  ..., -2.6837e-01,
     1.2549e-01,  9.6314e-01]],

  [[ 1.4472e-01,  1.5402e-01, -4.0394e-01,  ...,  5.9276e-01,
    -1.3165e-01, -1.2211e+00],
   [ 1.0750e+00, -1.0887e+00,  6.9819e-01,  ...,  6.7687e-01,
     1.4665e-01,  1.2630e+00],
   [ 8.3065e-01, -1.6116e-01,  5.3616e-02,  ...,  1.6874e+00,
    -9.2703e-01, -4.5544e-01],
   ...,
   [-2.3104e-01, -1.3227e+00, -9.4377e-01,  ...,  6.0654e-01,
    -9.6277e-01,  1.0423e+00],
   [ 5.4551e-01,  3.4231e-01,  3.7528e-01,  ...,  1.2852e+00,
     1.0643e+00, -1.3481e-03],
   [-4.1424e-01, -1.1618e+00,  3.8963e-02,  ..., -1.5521e-04,
    -1.2065e-01,  3.4900e-01]],

  [[ 1.3474e+00, -1.0317e-02, -6.0596e-01,  ..., -5.1366e-01,
    -1.3814e+00, -6.5122e-01],
   [ 8.0351e-01,  9.6109e-01,  1.4139e+00,  ...,  1.4535e+00,
     9.8968e-01,  4.2794e-01],
   [ 1.4144e-01, -6.9275e-01, -1.0767e+00,  ...,  2.9450e-01,
    -4.7638e-01,  1.6170e+00],
   ...,
   [ 1.8056e-01, -3.9629e-01, -1.1429e-03,  ...,  7.0036e-01,
    -8.7710e-01,  6.5044e-01],
   [ 1.2079e+00, -1.7680e+00,  1.5530e+00,  ...,  1.8241e+00,
     3.9575e-01, -5.3557e-02],
   [ 2.9455e-01, -2.7297e+00,  4.7883e-01,  ...,  3.2143e-01,
     1.0082e+00, -6.9631e-01]],

  ...,

  [[ 3.4586e-01, -1.0730e+00,  9.3181e-01,  ...,  5.9873e-01,
     2.6745e-01, -6.3752e-01],
   [ 1.9807e-01,  8.4666e-01,  1.7526e+00,  ..., -5.1485e-01,
     1.5930e+00, -1.3223e-01],
   [-2.0230e-01, -2.1122e+00, -5.2980e-01,  ...,  2.0098e+00,
    -1.2939e+00,  1.3258e+00],
   ...,
   [ 5.5298e-02, -1.6574e+00,  1.5176e+00,  ...,  5.3913e-01,
    -8.6454e-02, -7.3583e-01],
   [ 2.0102e-01, -1.4970e-01,  3.4424e-02,  ...,  6.0766e-01,
     9.4239e-01,  2.3055e-01],
   [ 3.8988e-01, -6.9052e-01,  1.0939e+00,  ..., -5.6343e-01,
    -8.9841e-01,  8.3196e-01]],

  [[-1.0901e+00, -1.2123e+00, -5.5091e-01,  ..., -1.1037e+00,
    -1.9915e+00,  1.9754e-01],
   [-1.3602e-01,  6.1435e-01,  1.5291e+00,  ...,  1.2577e+00,
     8.6335e-01,  1.5900e+00],
   [ 9.6869e-01, -7.9333e-01, -2.3843e-01,  ...,  3.0226e-01,
    -6.3827e-03,  8.3277e-01],
   ...,
   [ 1.4602e+00, -1.6949e+00,  5.7760e-01,  ...,  2.7545e+00,
     1.1102e-01,  3.2058e-01],
   [-3.7983e-02, -1.2483e-01,  1.6051e+00,  ...,  3.6712e-01,
     1.5149e+00, -2.0265e-01],
   [ 6.6216e-01, -9.5437e-01, -5.0103e-01,  ...,  5.7716e-01,
     7.4068e-01,  1.7688e-01]],

  [[ 1.0175e+00, -3.5308e-01,  7.2218e-02,  ...,  1.3013e+00,
    -2.2913e-01, -4.9519e-01],
   [-1.6743e-01,  5.3940e-01,  1.3644e+00,  ..., -4.3406e-01,
     2.0700e-02, -1.2070e-01],
   [ 4.9090e-02, -1.1509e+00, -8.9348e-01,  ..., -6.8726e-01,
    -6.2417e-01, -2.0275e-01],
   ...,
   [ 2.2913e-01, -1.8145e+00,  6.4851e-01,  ...,  2.1424e+00,
    -2.9331e-01,  5.9818e-01],
   [ 1.5226e+00, -1.0960e+00,  1.1153e+00,  ...,  1.7604e+00,
     5.2075e-01,  2.7665e-02],
   [-1.7653e+00, -2.9693e+00,  1.2386e+00,  ...,  7.7132e-01,
     2.3493e-01,  1.0911e+00]]]], grad_fn=<ViewBackward0>), permute_
     5.3991e-02, -1.1233e+00],
   [ 1.4472e-01,  1.5402e-01, -4.0394e-01,  ...,  5.9276e-01,
    -1.3165e-01, -1.2211e+00],
   [ 1.3474e+00, -1.0317e-02, -6.0596e-01,  ..., -5.1366e-01,
    1.3814e+00,  6.5133e-01]
```

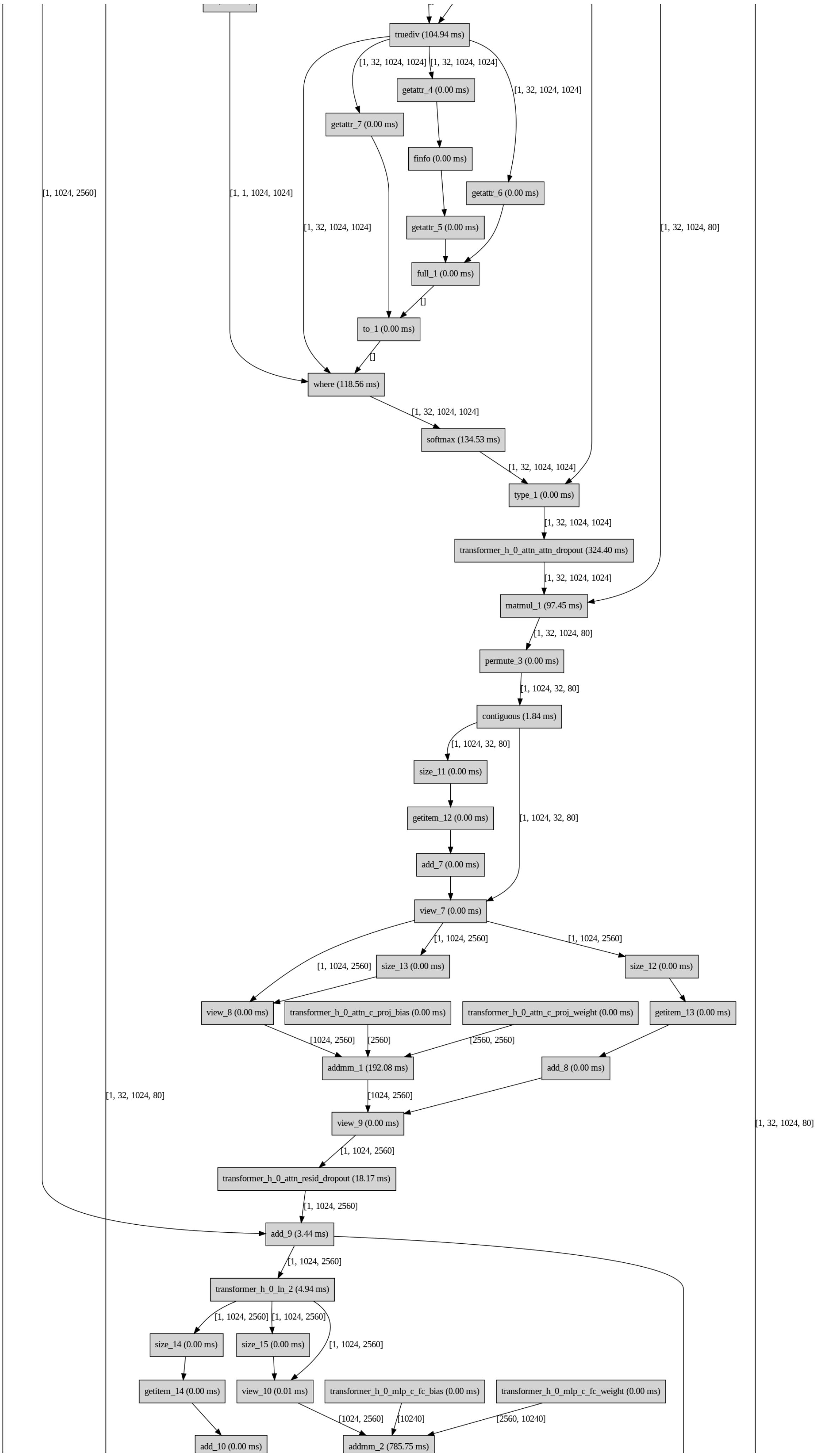
```
  -1.3814e+00, -6.5122e-01],
 ...,
 [ 3.4586e-01, -1.0730e+00,  9.3181e-01,  ...,  5.9873e-01,
   2.6745e-01, -6.3752e-01],
 [-1.0901e+00, -1.2123e+00, -5.5091e-01,  ..., -1.1037e+00,
  -1.9915e+00,  1.9754e-01],
 [ 1.0175e+00, -3.5308e-01,  7.2218e-02,  ...,  1.3013e+00,
  -2.2913e-01, -4.9519e-01]],

[[-3.1268e-01,  7.5064e-01,  1.3438e+00,  ..., -1.2461e-01,
   4.1225e-01,  1.0110e+00],
 [ 1.0750e+00, -1.0887e+00,  6.9819e-01,  ...,  6.7687e-01,
   1.4665e-01,  1.2630e+00],
 [ 8.0351e-01,  9.6109e-01,  1.4139e+00,  ...,  1.4535e+00,
   9.8968e-01,  4.2794e-01],
 ...,
 [ 1.9807e-01,  8.4666e-01,  1.7526e+00,  ..., -5.1485e-01,
   1.5930e+00, -1.3223e-01],
 [-1.3602e-01,  6.1435e-01,  1.5291e+00,  ...,  1.2577e+00,
   8.6335e-01,  1.5900e+00],
 [-1.6743e-01,  5.3940e-01,  1.3644e+00,  ..., -4.3406e-01,
   2.0700e-02, -1.2070e-01]],

[[ 5.3532e-01, -1.7978e-01, -9.8737e-01,  ...,  7.4715e-01,
  -4.8342e-01,  1.5643e+00],
 [ 8.3065e-01, -1.6116e-01,  5.3616e-02,  ...,  1.6874e+00,
  -9.2703e-01, -4.5544e-01],
 [ 1.4144e-01, -6.9275e-01, -1.0767e+00,  ...,  2.9450e-01,
  -4.7638e-01,  1.6170e+00],
 ...,
 [-2.0230e-01, -2.1122e+00, -5.2980e-01,  ...,  2.0098e+00,
  -1.2939e+00,  1.3258e+00],
 [ 9.6869e-01, -7.9333e-01, -2.3843e-01,  ...,  3.0226e-01,
  -6.3827e-03,  8.3277e-01],
 [ 4.9090e-02, -1.1509e+00, -8.9348e-01,  ..., -6.8726e-01,
  -6.2417e-01, -2.0275e-01]],


...,


[[-6.7070e-01, -9.6281e-01,  3.8247e-01,  ...,  1.4481e+00,
  -6.7513e-01,  7.8456e-01],
 [-2.3104e-01, -1.3227e+00, -9.4377e-01,  ...,  6.0654e-01,
  -9.6277e-01,  1.0423e+00],
 [ 1.8056e-01, -3.9629e-01, -1.1429e-03,  ...,  7.0036e-01,
  -8.7710e-01,  6.5044e-01],
 ...,
 [ 5.5298e-02, -1.6574e+00,  1.5176e+00,  ...,  5.3913e-01,
  -8.6454e-02, -7.3583e-01],
 [ 1.4602e+00, -1.6949e+00,  5.7760e-01,  ...,  2.7545e+00,
   1.1102e-01,  3.2058e-01],
 [ 2.2913e-01, -1.8145e+00,  6.4851e-01,  ...,  2.1424e+00,
  -2.9331e-01,  5.9818e-01]],

[[ 3.2699e-02, -1.7002e+00,  9.3860e-01,  ...,  1.0049e+00,
   1.3244e+00, -1.6203e+00],
 [ 5.4551e-01,  3.4231e-01,  3.7528e-01,  ...,  1.2852e+00,
   1.0643e+00, -1.3481e-03],
 [ 1.2079e+00, -1.7680e+00,  1.5530e+00,  ...,  1.8241e+00,
   2.0575e-01, -5.3557e-02]
   3.9575e-01, -5.3557e-02],
 ...,
 [ 2.0102e-01, -1.4970e-01,  3.4424e-02,  ...,  6.0766e-01,
   9.4239e-01,  2.3055e-01],
 [-3.7983e-02, -1.2483e-01,  1.6051e+00,  ...,  3.6712e-01,
   1.5149e+00, -2.0265e-01],
 [ 1.5226e+00, -1.0960e+00,  1.1153e+00,  ...,  1.7604e+00,
   5.2075e-01,  2.7665e-02]],

[[-2.1535e-01, -2.6749e+00,  2.0272e-01,  ..., -2.6837e-01,
   1.2549e-01,  9.6314e-01],
 [-4.1424e-01, -1.1618e+00,  3.8963e-02,  ..., -1.5521e-04,
  -1.2065e-01,  3.4900e-01],
 [ 2.9455e-01, -2.7297e+00,  4.7883e-01,  ...,  3.2143e-01,
   1.0082e+00, -6.9631e-01],
 ...,
 [ 3.8988e-01, -6.9052e-01,  1.0939e+00,  ..., -5.6343e-01,
  -8.9841e-01,  8.3196e-01],
 [ 6.6216e-01, -9.5437e-01, -5.0103e-01,  ...,  5.7716e-01,
   7.4068e-01,  1.7688e-01],
 [-1.7653e+00, -2.9693e+00,  1.2386e+00,  ...,  7.7132e-01,
   2.3493e-01,  1.0911e+00]]]], grad_fn=<PermuteBackward0>), trans
 [ 0.8811, -0.9153,  0.5432,  ..., -0.1132,  0.9024, -0.1774],
 [-0.1241,  0.2059,  0.0756,  ..., -0.0178,  0.1619,  1.1426],
 ...,
 [ 0.2266, -0.7287,  1.0897,  ..., -0.8767, -0.7393, -1.6697],
 [-0.7701, -1.3031, -1.7253,  ...,  0.9113, -1.1445,  0.7312],
 [-0.8997,  0.2322,  0.2846,  ...,  1.0387, -0.2149,  0.2506]],

[[-1.1550, -1.4128, -0.8195,  ..., -1.3198, -0.1504,  0.4139],
 [-0.6857,  0.3902, -1.1852,  ..., -0.0586, -0.2212,  0.7714],
 [-0.0548,  0.5208,  0.7765,  ...,  0.6382, -0.6004, -0.0446],
 ...,
 [ 0.7911,  1.5560,  1.4433,  ...,  0.1729,  0.8414,  1.4960],
 [-1.2452, -1.4400, -0.8974,  ..., -0.2085,  1.0264, -1.7593],
 [ 1.1871,  0.8671,  1.2275,  ...,  0.6523,  0.3428,  0.9587]],

[[ 0.9380, -0.2036,  0.5280,  ...,  0.3867, -0.2839, -0.2052],
 [-0.1100,  0.7863, -1.0059,  ...,  1.8113,  1.5350, -1.3002],
 [-1.4938,  0.6648,  1.2404,  ..., -0.2210, -0.5067, -0.2397],
 ...,
 [ 0.7169,  0.5309,  2.1611,  ..., -0.1997,  0.4349,  0.9969],
 [ 0.9322, -0.1505,  0.4663,  ...,  0.8334,  0.4209,  1.4980],
 [ 0.0979,  0.1302, -2.0759,  ..., -2.3713,  0.6408, -1.2450]],

...,

[[ 1.5353, -0.1049,  1.0090,  ...,  0.9555,  1.4842, -0.1303],
 [-0.6293, -1.6774, -0.6773,  ..., -0.3896, -0.3871, -2.1165],
 [ 0.1201,  0.6332,  0.5736,  ...,  1.3091,  1.0065,  1.7418],
 ...,
 [ 1.1989,  1.0535,  1.5946,  ...,  1.7802,  2.0689,  1.6911],
 [ 0.0139, -0.5428,  0.2422,  ...,  0.3060, -1.3078,  0.4163],
 [ 0.5774, -0.9433, -0.2608,  ...,  0.6750, -0.7658, -1.4566]],

[[-1.7977, -0.7844, -0.6200,  ..., -1.2726, -0.9818, -0.8389],
 [-0.0972,  0.0462,  0.4880,  ..., -0.3915,  0.4404, -0.8111],
 [ 0.2891, -1.1205, -1.5203,  ...,  0.6958, -0.4635, -0.7810],
   ...,
   [-0.4595,  0.4437,  0.4020,  ...,  0.4698,  0.2169, -0.4379],
   [-1.0992, -0.6814, -0.7304,  ..., -1.2266, -1.1535, -0.9150],
   [ 0.6587,  1.1193, -0.3302,  ..., -0.3855,  0.0350,  1.7493]],

  [[ 0.8641,  1.1099,  0.5323,  ...,  0.1825, -0.5885,  1.2332],
   [ 0.5925,  1.4441,  0.8540,  ..., -0.2630, -1.2048, -0.3148],
   [ 0.0250,  0.7351,  1.2001,  ..., -0.8271,  0.7633,  1.6390],
   ...,
   [ 0.9172,  1.5720,  0.8550,  ...,  0.5835,  2.2055,  1.5862],
   [ 0.8187, -0.0148,  0.7815,  ...,  0.7090,  2.1243,  1.8645],
   [ 0.0937,  0.3059, -0.1665,  ..., -1.3823,  1.1146, -0.2542]]]],
grad_fn=<TransposeBackward0>), matmul: tensor([[[[ 1.0430e+01,  4.67
     4.1251e+00,  2.2452e+00],
   [-7.6701e-01, -4.4139e+00,  4.5270e-01,  ...,  3.3855e+00,
    -1.0004e+01,  3.5770e+00],
   [-1.0984e+01, -1.2136e+01, -7.9908e+00,  ..., -1.4988e+00,
    -7.4073e+00, -5.4939e+00],
   ...,
   [ 2.0340e+00, -4.0041e-01,  1.1355e+01,  ...,  9.4114e+00,
     2.2624e+00, -3.0503e-01],
   [ 7.6295e+00,  3.0835e-01,  1.7212e+01,  ...,  7.0072e+00,
    -8.4236e-01,  4.6190e+00],
   [-9.1325e-01, -7.3764e+00,  1.7014e-01,  ..., -2.5813e+00,
    -5.6367e+00,  3.4909e-01]],

  [[ 5.5863e+00,  7.5378e+00,  5.1288e+00,  ..., -1.7925e+00,
     1.5287e+01, -2.0939e-02],
   [ 7.5467e+00,  1.2995e+00,  7.4965e+00,  ..., -1.8041e+00,
     1.9179e+01, -9.1187e-01],
   [ 1.9395e+00,  1.2405e+01,  1.1177e+01,  ...,  4.5255e+00,
     1.1279e+01, -4.8235e-01],
   ...,
   [ 8.1653e+00, -2.4434e+00,  1.4198e+01,  ...,  5.5239e+00,
     1.9794e+01, -1.8524e+01],
   [-4.0824e-01, -3.3011e+00,  4.8129e+00,  ...,  1.5234e+00,
    -2.1115e-01,  1.1105e+00],
   [ 2.4643e+00, -1.5341e+00,  7.8461e+00,  ..., -4.9817e+00,
     1.0209e+01, -5.6271e+00]],

  [[ 1.5906e+01,  1.2869e+01, -6.9656e+00,  ...,  1.1569e+01,
     7.8965e-01,  4.1757e-01],
   [-1.1141e-01,  1.1970e+00, -9.3608e+00,  ...,  1.0594e+01,
    -5.3976e+00, -3.5579e-03],
   [ 1.0445e+01,  1.1542e+01, -8.0955e-01,  ..., -1.0009e+00,
     9.8279e+00,  1.2675e+00],
   ...,
   [ 1.9157e+00,  6.0429e+00,  4.4252e+00,  ...,  1.2410e+01,
     4.0894e+00,  1.2655e+00],
   [ 1.4965e+01,  1.0058e+01, -6.5249e+00,  ..., -3.0971e-01,
    -4.3488e+00,  1.8811e-01],
   [ 1.0513e+01,  1.3291e+01, -2.1510e+00,  ...,  1.0560e+01,
     1.8814e+00,  3.6630e+00]],

  ...,

  [[ 4.2283e+00,  5.3545e+00,  4.3935e+00,  ..., -1.9658e+01,
    -4.0978e+00,  9.0045e+00],
   [ 2.8515e+00,  2.2677e-01,  3.0080e+00,  ..., -7.1774e+00,
     3.7882e+00,  1.0079e+01],
   [ 2.4718e-01,  3.5171e+00, -1.0701e+01,  ..., -1.2385e+01,
    -4.3520e+00,  3.6835e+00],
   ...,
   [ 7.2642e-01,  2.0913e+00, -4.8573e+00,  ..., -2.7290e+01,
    -4.5566e+00,  6.5560e-02],
   [ 1.5100e+00, -1.4625e+00, -4.0832e+00,  ..., -7.1641e+00,
    -4.3532e+00,  4.2141e+00],
   [ 1.1686e+00,  8.5506e+00,  1.0226e+01,  ..., -6.6008e+00,
     7.0959e+00,  1.6699e+01]],

  [[ 2.8508e+00,  4.5376e+00, -3.9154e+00,  ..., -8.3542e+00,
    -3.6045e-01,  4.9052e+00],
   [ 3.3203e+00,  7.4583e+00,  4.1673e-01,  ..., -3.6159e+00,
     9.5660e+00,  1.7008e+00],
   [ 5.4204e+00,  5.8717e+00,  4.3844e+00,  ..., -4.4120e+00,
    -4.5562e+00,  5.1776e+00],
   ...,
   [-4.1812e+00,  1.7936e-01, -4.1416e+00,  ..., -1.5494e+01,
    -1.2778e+00, -1.2007e+01],
   [ 1.1032e+01,  7.6278e+00,  5.4892e+00,  ..., -2.4466e-01,
     8.5342e+00,  9.8098e+00],
   [-5.2217e+00, -1.3395e+00, -8.4689e+00,  ..., -7.0843e+00,
    -1.0366e+00,  1.4809e+00]],

  [[ 7.8222e+00,  6.1463e+00,  3.6846e+00,  ...,  9.0359e+00,
     1.4237e+01,  4.5693e+00],
   [ 1.6756e+01,  1.6043e+01,  6.7309e+00,  ...,  1.5655e+01,
     1.2762e+01,  8.6012e+00],
   [-7.2196e+00, -4.5320e+00, -5.1647e+00,  ..., -2.2054e+00,
     1.3172e+00, -2.4509e+00],
   ...,
   [-2.7618e-01,  5.9328e-01, -5.6149e+00,  ...,  4.6296e+00,
    -1.5921e+01,  1.0148e+00],
   [-2.7612e+00, -1.0632e+01, -1.0232e+01,  ..., -2.6153e+00,
     6.0864e+00,  6.5763e+00],
   [-6.4058e-01, -1.7625e+00,  3.9127e+00,  ..., -1.1954e+00,
     9.8718e-01, -3.4104e+00]]]], grad_fn=<UnsafeViewBackward0>), si
     4.6120e-01,  2.5102e-01],
   [-8.5754e-02, -4.9349e-01,  5.0614e-02,  ...,  3.7850e-01,
    -1.1185e+00,  3.9992e-01],
   [-1.2281e+00, -1.3569e+00, -8.9340e-01,  ..., -1.6757e-01,
    -8.2816e-01, -6.1424e-01],
   ...,
   [ 2.2740e-01, -4.4767e-02,  1.2695e+00,  ...,  1.0522e+00,
     2.5294e-01, -3.4103e-02],
   [ 8.5300e-01,  3.4475e-02,  1.9244e+00,  ...,  7.8343e-01,
    -9.4179e-02,  5.1642e-01],
   [-1.0210e-01, -8.2470e-01,  1.9022e-02,  ..., -2.8859e-01,
    -6.3020e-01,  3.9030e-02]],

  [[ 6.2457e-01,  8.4275e-01,  5.7342e-01,  ..., -2.0040e-01,
     1.7091e+00, -2.3411e-03],
   [ 8.4374e-01,  1.4528e-01,  8.3813e-01,  ..., -2.0170e-01,
     2.1443e+00, -1.0195e-01],
   [ 2.1684e-01,  1.3869e+00,  1.2496e+00,  ...,  5.0596e-01,
     1.2610e+00, -5.3928e-02],
   ...,
   [ 9.1290e-01, -2.7318e-01,  1.5874e+00,  ...,  6.1759e-01,
     2.2131e+00, -2.0711e+00],
   [-4.5643e-02, -3.6907e-01,  5.3810e-01,  ...,  1.7032e-01,
    -2.3608e-02,  1.2416e-01],
   [ 2.7552e-01, -1.7152e-01,  8.7722e-01,  ..., -5.5697e-01,
     1.1414e+00, -6.2913e-01]],

  [[ 1.7783e+00,  1.4388e+00, -7.7878e-01,  ...,  1.2935e+00,
     8.8286e-02,  4.6685e-02],
   [-1.2456e-02,  1.3383e-01, -1.0466e+00,  ...,  1.1844e+00,
    -6.0347e-01, -3.9778e-04],
   [ 1.1678e+00,  1.2905e+00, -9.0510e-02,  ..., -1.1191e-01,
     1.0988e+00,  1.4171e-01],
   ...,
   [ 2.1418e-01,  6.7562e-01,  4.9475e-01,  ...,  1.3875e+00,
     4.5721e-01,  1.4149e-01],
   [ 1.6731e+00,  1.1245e+00, -7.2951e-01,  ..., -3.4626e-02,
    -4.8621e-01,  2.1031e-02],
   [ 1.1754e+00,  1.4860e+00, -2.4049e-01,  ...,  1.1806e+00,
     2.1035e-01,  4.0953e-01]],

  ...,

  [[ 4.7274e-01,  5.9865e-01,  4.9120e-01,  ..., -2.1978e+00,
    -4.5815e-01,  1.0067e+00],
   [ 3.1881e-01,  2.5353e-02,  3.3630e-01,  ..., -8.0246e-01,
     4.2354e-01,  1.1269e+00],
   [ 2.7636e-02,  3.9323e-01, -1.1964e+00,  ..., -1.3847e+00,
    -4.8657e-01,  4.1183e-01],
   ...,
   [ 8.1216e-02,  2.3382e-01, -5.4307e-01,  ..., -3.0511e+00,
    -5.0944e-01,  7.3299e-03],
   [ 1.6882e-01, -1.6351e-01, -4.5651e-01,  ..., -8.0097e-01,
    -4.8670e-01,  4.7115e-01],
   [ 1.3066e-01,  9.5599e-01,  1.1433e+00,  ..., -7.3799e-01,
     7.9335e-01,  1.8669e+00]],

  [[ 3.1872e-01,  5.0732e-01, -4.3776e-01,  ..., -9.3403e-01,
    -4.0299e-02,  5.4842e-01],
   [ 3.7122e-01,  8.3386e-01,  4.6592e-02,  ..., -4.0426e-01,
     1.0695e+00,  1.9015e-01],
   [ 6.0602e-01,  6.5647e-01,  4.9019e-01,  ..., -4.9328e-01,
    -5.0940e-01,  5.7888e-01],
   ...,
   [-4.6747e-01,  2.0053e-02, -4.6305e-01,  ..., -1.7323e+00,
    -1.4286e-01, -1.3424e+00],
   [ 1.2334e+00,  8.5281e-01,  6.1371e-01,  ..., -2.7353e-02,
     9.5415e-01,  1.0968e+00],
   [-5.8381e-01, -1.4976e-01, -9.4685e-01,  ..., -7.9205e-01,
    -1.1590e-01,  1.6557e-01]],

  [[ 8.7455e-01,  6.8718e-01,  4.1195e-01,  ...,  1.0102e+00,
     1.5917e+00,  5.1086e-01],
   [ 1.8733e+00,  1.7937e+00,  7.5253e-01,  ...,  1.7503e+00,
     1.4268e+00,  9.6165e-01],
   [-8.0718e-01, -5.0669e-01, -5.7743e-01,  ..., -2.4657e-01,
     1.4727e-01, -2.7402e-01],
   ...,
   [-3.0878e-02,  6.6331e-02, -6.2777e-01,  ...,  5.1760e-01,
    -1.7800e+00,  1.1346e-01],
   [-3.0871e-01, -1.1887e+00, -1.1440e+00,  ..., -2.9240e-01,
     6.8048e-01,  7.3525e-01],
   [-7.1619e-02, -1.9705e-01,  4.3745e-01,  ..., -1.3365e-01,
     1.1037e-01, -3.8129e-01]]]], grad_fn=<DivBackward0>), size_9: 1
   [1, 1, 0,  ..., 0, 0, 0],
   [1, 1, 1,  ..., 0, 0, 0],
   ...,
   [1, 1, 1,  ..., 1, 0, 0],
   [1, 1, 1,  ..., 1, 1, 0],
   [1, 1, 1,  ..., 1, 1, 1]]]], dtype=torch.uint8), sub: 0, getitem_
   [1, 1, 0,  ..., 0, 0, 0],
   [1, 1, 1,  ..., 0, 0, 0],
   ...,
   [1, 1, 1,  ..., 1, 0, 0],
   [1, 1, 1,  ..., 1, 1, 0],
   [1, 1, 1,  ..., 1, 1, 1]]]], dtype=torch.uint8), to: tensor([[[[
   [ True,  True, False,  ..., False, False, False],
   [ True,  True,  True,  ..., False, False, False],
   ...,
   [ True,  True,  True,  ...,  True, False, False],
   [ True,  True,  True,  ...,  True,  True, False],
   [ True,  True,  True,  ...,  True,  True,  True]]]]), getattr_4:
    -3.4028e+38, -3.4028e+38],
   [-8.5754e-02, -4.9349e-01, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [-1.2281e+00, -1.3569e+00, -8.9340e-01,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   ...,
   [ 2.2740e-01, -4.4767e-02,  1.2695e+00,  ...,  1.0522e+00,
    -3.4028e+38, -3.4028e+38],
   [ 8.5300e-01,  3.4475e-02,  1.9244e+00,  ...,  7.8343e-01,
    -9.4179e-02, -3.4028e+38],
   [-1.0210e-01, -8.2470e-01,  1.9022e-02,  ..., -2.8859e-01,
    -6.3020e-01,  3.9030e-02]],

  [[ 6.2457e-01, -3.4028e+38, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 8.4374e-01,  1.4528e-01, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 2.1684e-01,  1.3869e+00,  1.2496e+00,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   ...,
   [ 9.1290e-01, -2.7318e-01,  1.5874e+00,  ...,  6.1759e-01,
    -3.4028e+38, -3.4028e+38],
   [-4.5643e-02, -3.6907e-01,  5.3810e-01,  ...,  1.7032e-01,
    -2.3608e-02, -3.4028e+38],
   [ 2.7552e-01, -1.7152e-01,  8.7722e-01,  ..., -5.5697e-01,
     1.1414e+00, -6.2913e-01]],

  [[ 1.7783e+00, -3.4028e+38, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [-1.2456e-02,  1.3383e-01, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 1.1678e+00,  1.2905e+00, -9.0510e-02,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   ...,
   [ 2.1418e-01,  6.7562e-01,  4.9475e-01,  ...,  1.3875e+00,
    -3.4028e+38, -3.4028e+38],
   [ 1.6731e+00,  1.1245e+00, -7.2951e-01,  ..., -3.4626e-02,
    -4.8621e-01, -3.4028e+38],
   [ 1.1754e+00,  1.4860e+00, -2.4049e-01,  ...,  1.1806e+00,
     2.1035e-01,  4.0953e-01]],

  ...,

  [[ 4.7274e-01, -3.4028e+38, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 3.1881e-01,  2.5353e-02, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 2.7636e-02,  3.9323e-01, -1.1964e+00,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   ...,
   [ 8.1216e-02,  2.3382e-01, -5.4307e-01,  ..., -3.0511e+00,
    -3.4028e+38, -3.4028e+38],
   [ 1.6882e-01, -1.6351e-01, -4.5651e-01,  ..., -8.0097e-01,
    -4.8670e-01, -3.4028e+38],
   [ 1.3066e-01,  9.5599e-01,  1.1433e+00,  ..., -7.3799e-01,
     7.9335e-01,  1.8669e+00]],

  [[ 3.1872e-01, -3.4028e+38, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 3.7122e-01,  8.3386e-01, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 6.0602e-01,  6.5647e-01,  4.9019e-01,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   ...,
   [-4.6747e-01,  2.0053e-02, -4.6305e-01,  ..., -1.7323e+00,
    -3.4028e+38, -3.4028e+38],
   [ 1.2334e+00,  8.5281e-01,  6.1371e-01,  ..., -2.7353e-02,
     9.5415e-01, -3.4028e+38],
   [-5.8381e-01, -1.4976e-01, -9.4685e-01,  ..., -7.9205e-01,
    -1.1590e-01,  1.6557e-01]],

  [[ 8.7455e-01, -3.4028e+38, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [ 1.8733e+00,  1.7937e+00, -3.4028e+38,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   [-8.0718e-01, -5.0669e-01, -5.7743e-01,  ..., -3.4028e+38,
    -3.4028e+38, -3.4028e+38],
   ...,
   [-3.0878e-02,  6.6331e-02, -6.2777e-01,  ...,  5.1760e-01,
    -3.4028e+38, -3.4028e+38],
   [-3.0871e-01, -1.1887e+00, -1.1440e+00,  ..., -2.9240e-01,
     6.8048e-01, -3.4028e+38],
   [-7.1619e-02, -1.9705e-01,  4.3745e-01,  ..., -1.3365e-01,
     1.1037e-01, -3.8129e-01]]]], grad_fn=<WhereBackward0>), softmax
    0.0000e+00, 0.0000e+00],
   [6.0054e-01, 3.9946e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [3.0519e-01, 2.6831e-01, 4.2650e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [6.3499e-04, 4.8369e-04, 1.8003e-03,  ..., 1.4487e-03,
    0.0000e+00, 0.0000e+00],
    0.0000e+00, 0.0000e+00],
   [1.0478e-03, 4.6214e-04, 3.0588e-03,  ..., 9.7734e-04,
    4.0635e-04, 0.0000e+00],
   [8.5923e-04, 4.1715e-04, 9.6987e-04,  ..., 7.1305e-04,
    5.0671e-04, 9.8947e-04]],

  [[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [6.6785e-01, 3.3215e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [1.4223e-01, 4.5828e-01, 3.9949e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [1.0453e-03, 3.1925e-04, 2.0520e-03,  ..., 7.7802e-04,
    0.0000e+00, 0.0000e+00],
   [4.3650e-04, 3.1588e-04, 7.8253e-04,  ..., 5.4172e-04,
    4.4622e-04, 0.0000e+00],
   [9.8641e-04, 6.3083e-04, 1.8004e-03,  ..., 4.2905e-04,
    2.3448e-03, 3.9918e-04]],

  [[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [4.6349e-01, 5.3651e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [4.1413e-01, 4.6820e-01, 1.1767e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [6.3110e-04, 1.0011e-03, 8.3551e-04,  ..., 2.0401e-03,
    0.0000e+00, 0.0000e+00],
   [2.1595e-03, 1.2476e-03, 1.9539e-04,  ..., 3.9146e-04,
    2.4921e-04, 0.0000e+00],
   [4.8377e-04, 6.5998e-04, 1.1742e-04,  ..., 4.8630e-04,
    1.8430e-04, 2.2493e-04]],


  ...,


  [[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [5.7284e-01, 4.2716e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [3.6558e-01, 5.2693e-01, 1.0750e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [2.3302e-03, 2.7144e-03, 1.2482e-03,  ..., 1.0164e-04,
    0.0000e+00, 0.0000e+00],
   [1.6066e-03, 1.1523e-03, 8.5966e-04,  ..., 6.0916e-04,
    8.3409e-04, 0.0000e+00],
   [1.0623e-03, 2.4249e-03, 2.9243e-03,  ..., 4.4566e-04,
    2.0609e-03, 6.0300e-03]],

  [[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [3.8636e-01, 6.1364e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [3.3986e-01, 3.5745e-01, 3.0269e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [6.8848e-04, 1.1210e-03, 6.9153e-04,  ..., 1.9435e-04,
    0.0000e+00, 0.0000e+00],
   [8.6790e-04, 5.9318e-04, 4.6703e-04,  ..., 2.4600e-04,
    6.5644e-04, 0.0000e+00],
   [5.0960e-04, 7.8656e-04, 3.5445e-04,  ..., 4.1380e-04,
    8.1365e-04, 1.0781e-03]],

  [[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [5.1990e-01, 4.8010e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [2.7710e-01, 3.7423e-01, 3.4867e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [8.3461e-04, 9.1981e-04, 4.5947e-04,  ..., 1.4444e-03,
    0.0000e+00, 0.0000e+00],
   [6.1185e-04, 2.5379e-04, 2.6539e-04,  ..., 6.2190e-04,
    1.6453e-03, 0.0000e+00],
   [9.2261e-04, 8.1384e-04, 1.5350e-03,  ..., 8.6712e-04,
    1.1068e-03, 6.7690e-04]]]], grad_fn=<SoftmaxBackward0>), getattr
    0.0000e+00, 0.0000e+00],
   [6.0054e-01, 3.9946e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [3.0519e-01, 2.6831e-01, 4.2650e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [6.3499e-04, 4.8369e-04, 1.8003e-03,  ..., 1.4487e-03,
    0.0000e+00, 0.0000e+00],
   [1.0478e-03, 4.6214e-04, 3.0588e-03,  ..., 9.7734e-04,
    4.0635e-04, 0.0000e+00],
   [8.5923e-04, 4.1715e-04, 9.6987e-04,  ..., 7.1305e-04,
    5.0671e-04, 9.8947e-04]],

  [[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [6.6785e-01, 3.3215e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [1.4223e-01, 4.5828e-01, 3.9949e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [1.0453e-03, 3.1925e-04, 2.0520e-03,  ..., 7.7802e-04,
    0.0000e+00, 0.0000e+00],
   [4.3650e-04, 3.1588e-04, 7.8253e-04,  ..., 5.4172e-04,
    4.4622e-04, 0.0000e+00],
   [9.8641e-04, 6.3083e-04, 1.8004e-03,  ..., 4.2905e-04,
    2.3448e-03, 3.9918e-04]],

  [[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [4.6349e-01, 5.3651e-01, 0.0000e+00,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   [4.1413e-01, 4.6820e-01, 1.1767e-01,  ..., 0.0000e+00,
    0.0000e+00, 0.0000e+00],
   ...,
   [6.3110e-04, 1.0011e-03, 8.3551e-04,  ..., 2.0401e-03,
    0.0000e+00, 0.0000e+00],
   [2.1595e-03, 1.2476e-03, 1.9539e-04,  ..., 3.9146e-04,
    2.4921e-04, 0.0000e+00],
   [4.8377e-04, 6.5998e-04, 1.1742e-04,  ..., 4.8630e-04,
    1.8430e-04, 2.3403e-04]]
```

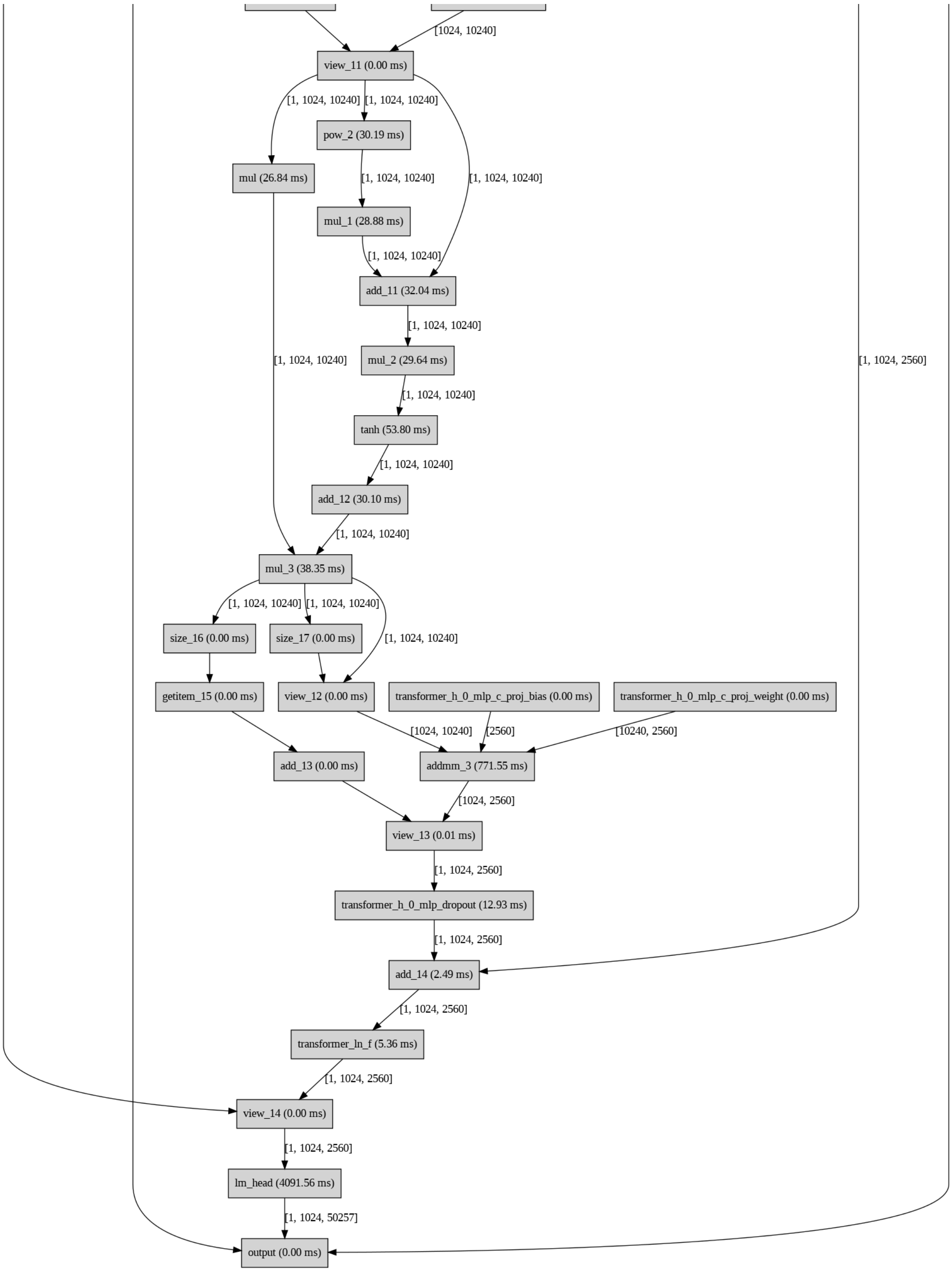
```
  1.8430e-04, 2.2493e-04]],

...,

[[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [5.7284e-01, 4.2716e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [3.6558e-01, 5.2693e-01, 1.0750e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [2.3302e-03, 2.7144e-03, 1.2482e-03,  ..., 1.0164e-04,
  0.0000e+00, 0.0000e+00],
 [1.6066e-03, 1.1523e-03, 8.5966e-04,  ..., 6.0916e-04,
  8.3409e-04, 0.0000e+00],
 [1.0623e-03, 2.4249e-03, 2.9243e-03,  ..., 4.4566e-04,
  2.0609e-03, 6.0300e-03]],

[[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [3.8636e-01, 6.1364e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [3.3986e-01, 3.5745e-01, 3.0269e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [6.8848e-04, 1.1210e-03, 6.9153e-04,  ..., 1.9435e-04,
  0.0000e+00, 0.0000e+00],
 [8.6790e-04, 5.9318e-04, 4.6703e-04,  ..., 2.4600e-04,
  6.5644e-04, 0.0000e+00],
 [5.0960e-04, 7.8656e-04, 3.5445e-04,  ..., 4.1380e-04,
  8.1365e-04, 1.0781e-03]],

[[1.0000e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [5.1990e-01, 4.8010e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [2.7710e-01, 3.7423e-01, 3.4867e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [8.3461e-04, 9.1981e-04, 4.5947e-04,  ..., 1.4444e-03,
  0.0000e+00, 0.0000e+00],
 [6.1185e-04, 2.5379e-04, 2.6539e-04,  ..., 6.2190e-04,
  1.6453e-03, 0.0000e+00],
 [9.2261e-04, 8.1384e-04, 1.5350e-03,  ..., 8.6712e-04,
  1.1068e-03, 6.7690e-04]]]], grad_fn=<SoftmaxBackward0>), transfo
  0.0000e+00, 0.0000e+00],
 [6.6727e-01, 4.4384e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [3.3910e-01, 2.9813e-01, 4.7389e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [7.0555e-04, 5.3743e-04, 2.0003e-03,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [1.1642e-03, 5.1349e-04, 3.3987e-03,  ..., 1.0859e-03,
  0.0000e+00, 0.0000e+00],
 [9.5470e-04, 4.6350e-04, 1.0776e-03,  ..., 7.9227e-04,
  5.6301e-04, 1.0994e-03]],

[[1.1111e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
[[1.1111e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [7.4205e-01, 3.6906e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [1.5803e-01, 5.0920e-01, 4.4387e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [0.0000e+00, 3.5472e-04, 2.2800e-03,  ..., 8.6447e-04,
  0.0000e+00, 0.0000e+00],
 [4.8500e-04, 3.5098e-04, 8.6948e-04,  ..., 6.0191e-04,
  4.9580e-04, 0.0000e+00],
 [1.0960e-03, 7.0092e-04, 2.0005e-03,  ..., 4.7673e-04,
  0.0000e+00, 4.4354e-04]],

[[1.1111e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [5.1499e-01, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [4.6014e-01, 5.2022e-01, 1.3075e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [7.0122e-04, 1.1124e-03, 9.2834e-04,  ..., 2.2668e-03,
  0.0000e+00, 0.0000e+00],
 [2.3994e-03, 1.3862e-03, 2.1710e-04,  ..., 4.3495e-04,
  2.7690e-04, 0.0000e+00],
 [5.3753e-04, 7.3331e-04, 1.3047e-04,  ..., 5.4034e-04,
  2.0478e-04, 2.4992e-04]],

...,

[[1.1111e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [0.0000e+00, 4.7462e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [0.0000e+00, 5.8547e-01, 1.1944e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [2.5892e-03, 0.0000e+00, 1.3869e-03,  ..., 1.1293e-04,
  0.0000e+00, 0.0000e+00],
 [1.7851e-03, 1.2804e-03, 9.5517e-04,  ..., 6.7685e-04,
  9.2677e-04, 0.0000e+00],
 [1.1803e-03, 2.6943e-03, 3.2493e-03,  ..., 4.9518e-04,
  2.2899e-03, 0.0000e+00]],

[[1.1111e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [4.2929e-01, 6.8182e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [3.7762e-01, 3.9717e-01, 3.3632e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [7.6498e-04, 1.2456e-03, 7.6837e-04,  ..., 2.1594e-04,
  0.0000e+00, 0.0000e+00],
 [9.6433e-04, 6.5909e-04, 5.1892e-04,  ..., 2.7333e-04,
  7.2938e-04, 0.0000e+00],
 [5.6622e-04, 8.7395e-04, 3.9383e-04,  ..., 4.5977e-04,
  9.0405e-04, 1.1979e-03]],

[[1.1111e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
[[1.1111e+00, 0.0000e+00, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [5.7767e-01, 5.3344e-01, 0.0000e+00,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 [3.0789e-01, 4.1581e-01, 3.8741e-01,  ..., 0.0000e+00,
  0.0000e+00, 0.0000e+00],
 ...,
 [9.2734e-04, 1.0220e-03, 5.1052e-04,  ..., 1.6049e-03,
  0.0000e+00, 0.0000e+00],
 [6.7983e-04, 2.8199e-04, 0.0000e+00,  ..., 6.9101e-04,
  1.8281e-03, 0.0000e+00],
 [1.0251e-03, 9.0427e-04, 1.7055e-03,  ..., 9.6346e-04,
  0.0000e+00, 7.5211e-04]]]], grad_fn=<MulBackward0>), matmul_1: t
 [ 0.4277,  0.3429,  0.2743,  ...,  0.5247, -0.0224, -1.2915],
 [ 0.8664,  0.1805, -0.1771,  ...,  0.0663, -0.6756, -1.0535],
 ...,
 [-0.2023, -0.1170,  0.0833,  ..., -0.0799, -0.6539, -0.6087],
 [-0.1736, -0.1207,  0.0532,  ..., -0.1172, -0.6309, -0.6744],
 [-0.1923, -0.1259,  0.0798,  ..., -0.0971, -0.6337, -0.6659]],

[[-0.3474,  0.8340,  1.4931,  ..., -0.1385,  0.4581,  1.1233],
 [ 0.1647,  0.1552,  1.2549,  ...,  0.1573,  0.3600,  1.2163],
 [ 0.8546, -0.0091,  1.1955,  ...,  0.9701,  0.5791,  0.9929],
 ...,
 [ 0.7826,  0.2138,  1.2433,  ..., -0.1273,  0.3371,  0.8644],
 [ 0.7615,  0.2055,  1.1951,  ..., -0.1102,  0.3089,  0.8784],
 [ 0.7613,  0.2119,  1.2075,  ..., -0.0870,  0.3660,  0.8562]],

[[ 0.5948, -0.1998, -1.0971,  ...,  0.8302, -0.5371,  1.7381],
 [ 0.2757, -0.0926, -0.5085,  ...,  0.3848, -0.2490,  0.8056],
 [ 0.6969, -0.2571, -0.5672,  ...,  1.2601, -0.7670,  0.6943],
 ...,
 [ 0.1452, -0.4147, -1.0127,  ...,  0.5365, -0.3885,  0.6983],
 [ 0.1689, -0.3585, -1.0242,  ...,  0.5535, -0.4173,  0.7276],
 [ 0.1675, -0.3300, -1.0409,  ...,  0.5176, -0.3676,  0.6784]],

...,

[[-0.7452, -1.0698,  0.4250,  ...,  1.6089, -0.7501,  0.8717],
 [-0.1097, -0.6278, -0.4479,  ...,  0.2879, -0.4570,  0.4947],
 [-0.1137, -0.8218, -0.5527,  ...,  0.4388, -0.6684,  0.6879],
 ...,
 [ 0.4239, -1.2654, -0.3864,  ...,  0.9375, -0.2147,  0.5535],
 [ 0.4094, -1.2832, -0.3667,  ...,  0.9196, -0.2319,  0.5110],
 [ 0.4249, -1.3403, -0.4202,  ...,  0.9643, -0.2362,  0.5460]],

[[ 0.0363, -1.8892,  1.0429,  ...,  1.1165,  1.4715, -1.8004],
 [ 0.3860, -0.4965,  0.6588,  ...,  1.3076,  1.2942, -0.6965],
 [ 0.6352, -1.1007,  1.0258,  ...,  1.5034,  1.0559, -0.6304],
 ...,
 [ 0.0457, -0.5143,  0.7223,  ...,  1.4657,  0.8046, -0.0067],
 [ 0.0658, -0.4800,  0.6999,  ...,  1.4708,  0.7521, -0.0510],
 [ 0.0502, -0.4834,  0.7533,  ...,  1.4894,  0.7789, -0.0379]],

[[-0.2393, -2.9722,  0.2252,  ..., -0.2982,  0.1394,  1.0702],
 [-0.3454, -2.1650,  0.1379,  ..., -0.1551,  0.0081,  0.7425],
 [-0.1244, -2.3642,  0.2641,  ...,  0.0418,  0.3791,  0.1719],
 ...,
 [ 0.0244, -1.6727,  0.0908,  ...,  0.3804,  0.3102,  0.4138]
```

```
   [ 0.0244, -1.6727,  0.0908,  ..., -0.3804,  0.2192,  0.4138],
   [ 0.0608, -1.7108,  0.1040,  ..., -0.4188,  0.2411,  0.4586],
   [ 0.0065, -1.6081,  0.0938,  ..., -0.4156,  0.2129,  0.4586]]]],
grad_fn=<UnsafeViewBackward0>), permute_3: tensor([[[[ 0.6053,  0.45
   [-0.3474,  0.8340,  1.4931,  ..., -0.1385,  0.4581,  1.1233],
   [ 0.5948, -0.1998, -1.0971,  ...,  0.8302, -0.5371,  1.7381],
   ...,
   [-0.7452, -1.0698,  0.4250,  ...,  1.6089, -0.7501,  0.8717],
   [ 0.0363, -1.8892,  1.0429,  ...,  1.1165,  1.4715, -1.8004],
   [-0.2393, -2.9722,  0.2252,  ..., -0.2982,  0.1394,  1.0702]],

  [[ 0.4277,  0.3429,  0.2743,  ...,  0.5247, -0.0224, -1.2915],
   [ 0.1647,  0.1552,  1.2549,  ...,  0.1573,  0.3600,  1.2163],
   [ 0.2757, -0.0926, -0.5085,  ...,  0.3848, -0.2490,  0.8056],
   ...,
   [-0.1097, -0.6278, -0.4479,  ...,  0.2879, -0.4570,  0.4947],
   [ 0.3860, -0.4965,  0.6588,  ...,  1.3076,  1.2942, -0.6965],
   [-0.3454, -2.1650,  0.1379,  ..., -0.1551,  0.0081,  0.7425]],

  [[ 0.8664,  0.1805, -0.1771,  ...,  0.0663, -0.6756, -1.0535],
   [ 0.8546, -0.0091,  1.1955,  ...,  0.9701,  0.5791,  0.9929],
   [ 0.6969, -0.2571, -0.5672,  ...,  1.2601, -0.7670,  0.6943],
   ...,
   [-0.1137, -0.8218, -0.5527,  ...,  0.4388, -0.6684,  0.6879],
   [ 0.6352, -1.1007,  1.0258,  ...,  1.5034,  1.0559, -0.6304],
   [-0.1244, -2.3642,  0.2641,  ...,  0.0418,  0.3791,  0.1719]],

  ...,

  [[-0.2023, -0.1170,  0.0833,  ..., -0.0799, -0.6539, -0.6087],
   [ 0.7826,  0.2138,  1.2433,  ..., -0.1273,  0.3371,  0.8644],
   [ 0.1452, -0.4147, -1.0127,  ...,  0.5365, -0.3885,  0.6983],
   ...,
   [ 0.4239, -1.2654, -0.3864,  ...,  0.9375, -0.2147,  0.5535],
   [ 0.0457, -0.5143,  0.7223,  ...,  1.4657,  0.8046, -0.0067],
   [ 0.0244, -1.6727,  0.0908,  ..., -0.3804,  0.2192,  0.4138]],

  [[-0.1736, -0.1207,  0.0532,  ..., -0.1172, -0.6309, -0.6744],
   [ 0.7615,  0.2055,  1.1951,  ..., -0.1102,  0.3089,  0.8784],
   [ 0.1689, -0.3585, -1.0242,  ...,  0.5535, -0.4173,  0.7276],
   ...,
   [ 0.4094, -1.2832, -0.3667,  ...,  0.9196, -0.2319,  0.5110],
   [ 0.0658, -0.4800,  0.6999,  ...,  1.4708,  0.7521, -0.0510],
   [ 0.0608, -1.7108,  0.1040,  ..., -0.4188,  0.2411,  0.4586]],

  [[-0.1923, -0.1259,  0.0798,  ..., -0.0971, -0.6337, -0.6659],
   [ 0.7613,  0.2119,  1.2075,  ..., -0.0870,  0.3660,  0.8562],
   [ 0.1675, -0.3300, -1.0409,  ...,  0.5176, -0.3676,  0.6784],
   ...,
   [ 0.4249, -1.3403, -0.4202,  ...,  0.9643, -0.2362,  0.5460],
   [ 0.0502, -0.4834,  0.7533,  ...,  1.4894,  0.7789, -0.0379],
   [ 0.0065, -1.6081,  0.0938,  ..., -0.4156,  0.2129,  0.4586]]]],
grad_fn=<PermuteBackward0>), contiguous: tensor([[[[ 0.6053,  0.4571
   [-0.3474,  0.8340,  1.4931,  ..., -0.1385,  0.4581,  1.1233],
   [ 0.5948, -0.1998, -1.0971,  ...,  0.8302, -0.5371,  1.7381],
   ...,
   [-0.7452, -1.0698,  0.4250,  ...,  1.6089, -0.7501,  0.8717],
   [ 0.0363, -1.8892,  1.0429,  ...,  1.1165,  1.4715, -1.8004],
   [-0.2393, -2.9722,  0.2252,  ..., -0.2982,  0.1394,  1.0702]],
   [-0.2393, -2.9722,  0.2252,  ..., -0.2982,  0.1394,  1.0702]],

  [[ 0.4277,  0.3429,  0.2743,  ...,  0.5247, -0.0224, -1.2915],
   [ 0.1647,  0.1552,  1.2549,  ...,  0.1573,  0.3600,  1.2163],
   [ 0.2757, -0.0926, -0.5085,  ...,  0.3848, -0.2490,  0.8056],
   ...,
   [-0.1097, -0.6278, -0.4479,  ...,  0.2879, -0.4570,  0.4947],
   [ 0.3860, -0.4965,  0.6588,  ...,  1.3076,  1.2942, -0.6965],
   [-0.3454, -2.1650,  0.1379,  ..., -0.1551,  0.0081,  0.7425]],

  [[ 0.8664,  0.1805, -0.1771,  ...,  0.0663, -0.6756, -1.0535],
   [ 0.8546, -0.0091,  1.1955,  ...,  0.9701,  0.5791,  0.9929],
   [ 0.6969, -0.2571, -0.5672,  ...,  1.2601, -0.7670,  0.6943],
   ...,
   [-0.1137, -0.8218, -0.5527,  ...,  0.4388, -0.6684,  0.6879],
   [ 0.6352, -1.1007,  1.0258,  ...,  1.5034,  1.0559, -0.6304],
   [-0.1244, -2.3642,  0.2641,  ...,  0.0418,  0.3791,  0.1719]],

  ...,

  [[-0.2023, -0.1170,  0.0833,  ..., -0.0799, -0.6539, -0.6087],
   [ 0.7826,  0.2138,  1.2433,  ..., -0.1273,  0.3371,  0.8644],
   [ 0.1452, -0.4147, -1.0127,  ...,  0.5365, -0.3885,  0.6983],
   ...,
   [ 0.4239, -1.2654, -0.3864,  ...,  0.9375, -0.2147,  0.5535],
   [ 0.0457, -0.5143,  0.7223,  ...,  1.4657,  0.8046, -0.0067],
   [ 0.0244, -1.6727,  0.0908,  ..., -0.3804,  0.2192,  0.4138]],

  [[-0.1736, -0.1207,  0.0532,  ..., -0.1172, -0.6309, -0.6744],
   [ 0.7615,  0.2055,  1.1951,  ..., -0.1102,  0.3089,  0.8784],
   [ 0.1689, -0.3585, -1.0242,  ...,  0.5535, -0.4173,  0.7276],
   ...,
   [ 0.4094, -1.2832, -0.3667,  ...,  0.9196, -0.2319,  0.5110],
   [ 0.0658, -0.4800,  0.6999,  ...,  1.4708,  0.7521, -0.0510],
   [ 0.0608, -1.7108,  0.1040,  ..., -0.4188,  0.2411,  0.4586]],

  [[-0.1923, -0.1259,  0.0798,  ..., -0.0971, -0.6337, -0.6659],
   [ 0.7613,  0.2119,  1.2075,  ..., -0.0870,  0.3660,  0.8562],
   [ 0.1675, -0.3300, -1.0409,  ...,  0.5176, -0.3676,  0.6784],
   ...,
   [ 0.4249, -1.3403, -0.4202,  ...,  0.9643, -0.2362,  0.5460],
   [ 0.0502, -0.4834,  0.7533,  ...,  1.4894,  0.7789, -0.0379],
   [ 0.0065, -1.6081,  0.0938,  ..., -0.4156,  0.2129,  0.4586]]]],
grad_fn=<CloneBackward0>), size_11: torch.Size([1, 1024, 32, 80]), g
  [ 0.4277,  0.3429,  0.2743,  ..., -0.1551,  0.0081,  0.7425],
  [ 0.8664,  0.1805, -0.1771,  ...,  0.0418,  0.3791,  0.1719],
  ...,
  [-0.2023, -0.1170,  0.0833,  ..., -0.3804,  0.2192,  0.4138],
  [-0.1736, -0.1207,  0.0532,  ..., -0.4188,  0.2411,  0.4586],
  [-0.1923, -0.1259,  0.0798,  ..., -0.4156,  0.2129,  0.4586]]],
grad_fn=<ViewBackward0>), size_12: torch.Size([1, 1024, 2560]), geti
 [ 0.4277,  0.3429,  0.2743,  ..., -0.1551,  0.0081,  0.7425],
 [ 0.8664,  0.1805, -0.1771,  ...,  0.0418,  0.3791,  0.1719],
 ...,
 [-0.2023, -0.1170,  0.0833,  ..., -0.3804,  0.2192,  0.4138],
 [-0.1736, -0.1207,  0.0532,  ..., -0.4188,  0.2411,  0.4586],
 [-0.1923, -0.1259,  0.0798,  ..., -0.4156,  0.2129,  0.4586]],
grad_fn=<ViewBackward0>), transformer_h_0_attn_c_proj_weight: tensor
 [-0.0147, -0.0144,  0.0186,  ..., -0.0061,  0.0135,  0.0176],
 [ 0.0147, -0.0144,  0.0186,  ..., -0.0061, -0.0135, -0.0176],
 [-0.0326,  0.0170, -0.0065,  ..., -0.0038,  0.0025,  0.0330],
 ...,
 [-0.0344, -0.0199,  0.0153,  ..., -0.0183, -0.0159, -0.0020],
 [-0.0060, -0.0005,  0.0059,  ..., -0.0068,  0.0180,  0.0233],
 [ 0.0297, -0.0124,  0.0067,  ...,  0.0089, -0.0263,  0.0058]]), add
 [ 0.2071,  0.5983,  0.0350,  ...,  0.4182,  0.6589, -0.4245],
 [ 0.6145,  0.5085, -0.5767,  ..., -0.2871,  1.0340,  0.1270],
 ...,
 [ 0.4360,  0.2205, -0.6687,  ..., -0.3988,  0.7918, -0.3463],
 [ 0.4293,  0.1773, -0.6789,  ..., -0.4150,  0.7921, -0.3534],
 [ 0.4610,  0.2209, -0.6895,  ..., -0.3759,  0.8103, -0.3680]],
grad_fn=<AddmmBackward0>), view_9: tensor([[[ 0.5176,  0.7158,  0.86
  [ 0.2071,  0.5983,  0.0350,  ...,  0.4182,  0.6589, -0.4245],
  [ 0.6145,  0.5085, -0.5767,  ..., -0.2871,  1.0340,  0.1270],
  ...,
  [ 0.4360,  0.2205, -0.6687,  ..., -0.3988,  0.7918, -0.3463],
  [ 0.4293,  0.1773, -0.6789,  ..., -0.4150,  0.7921, -0.3534],
  [ 0.4610,  0.2209, -0.6895,  ..., -0.3759,  0.8103, -0.3680]]],
grad_fn=<ViewBackward0>), transformer_h_0_attn_resid_dropout: tensor
  [ 0.2301,  0.6648,  0.0388,  ...,  0.4647,  0.7321, -0.4716],
  [ 0.6828,  0.5650, -0.6408,  ..., -0.0000,  1.1489,  0.0000],
  ...,
  [ 0.4844,  0.2450, -0.7430,  ..., -0.4431,  0.8798, -0.3848],
  [ 0.4770,  0.1970, -0.7544,  ..., -0.4611,  0.8801, -0.3926],
  [ 0.5122,  0.0000, -0.7661,  ..., -0.4176,  0.0000, -0.4089]]],
grad_fn=<MulBackward0>), add_9: tensor([[[ 0.6074,  0.8005,  0.9852,
  [ 0.2301,  0.6352,  0.0131,  ...,  0.4838,  0.7474, -0.4893],
  [ 0.7017,  0.5057, -0.6122,  ..., -0.0089,  1.1697,  0.0000],
  ...,
  [ 0.4838,  0.2021, -0.7034,  ..., -0.4508,  0.8916, -0.4399],
  [ 0.4801,  0.1672, -0.7934,  ..., -0.4822,  0.8677, -0.3999],
  [ 0.5309, -0.0608, -0.7760,  ..., -0.3705, -0.0052, -0.4417]]],
grad_fn=<AddBackward0>), transformer_h_0_ln_2: tensor([[[ 7.6818e-01
    8.9789e-01, -5.3465e-02],
  [ 3.3221e-01,  9.1876e-01,  1.7940e-02,  ...,  6.9946e-01,
    1.0811e+00, -7.0948e-01],
  [ 1.0707e+00,  7.7512e-01, -9.1016e-01,  ..., -5.7635e-04,
    1.7761e+00,  1.2798e-02],
  ...,
  [ 9.2372e-01,  3.8169e-01, -1.3609e+00,  ..., -8.7479e-01,
    1.7085e+00, -8.5374e-01],
  [ 9.3734e-01,  3.2845e-01, -1.5409e+00,  ..., -9.3523e-01,
    1.6916e+00, -7.7510e-01],
  [ 1.0397e+00, -1.1048e-01, -1.5006e+00,  ..., -7.1251e-01,
   -2.3505e-03, -8.5077e-01]]], grad_fn=<NativeLayerNormBackward0>),
   8.9789e-01, -5.3465e-02],
 [ 3.3221e-01,  9.1876e-01,  1.7940e-02,  ...,  6.9946e-01,
   1.0811e+00, -7.0948e-01],
 [ 1.0707e+00,  7.7512e-01, -9.1016e-01,  ..., -5.7635e-04,
   1.7761e+00,  1.2798e-02],
 ...,
 [ 9.2372e-01,  3.8169e-01, -1.3609e+00,  ..., -8.7479e-01,
   1.7085e+00, -8.5374e-01],
 [ 9.3734e-01,  3.2845e-01, -1.5409e+00,  ..., -9.3523e-01,
   1.6916e+00, -7.7510e-01],
 [ 1.0397e+00, -1.1048e-01, -1.5006e+00,  ..., -7.1251e-01,
  -2.3505e-03, -8.5077e-01]], grad_fn=<ViewBackward0>), transformer_
 [ 0.0128,  0.0525,  0.0116,  ...,  0.0020,  0.0212,  0.0076]
 [-0.0138, -0.0525,  0.0116,  ..., -0.0020,  0.0212,  0.0076],
 [ 0.0152, -0.0384, -0.0191,  ...,  0.0015, -0.0177, -0.0153],
 ...,
 [-0.0081, -0.0278,  0.0184,  ...,  0.0412, -0.0199,  0.0130],
 [-0.0073, -0.0078,  0.0139,  ..., -0.0191,  0.0340,  0.0254],
 [-0.0109,  0.0057, -0.0169,  ...,  0.0317, -0.0119,  0.0022]]), add
 [ 0.4070,  1.9528, -1.1921,  ...,  0.7105, -0.8874, -0.2245],
 [-1.3988,  2.6029, -2.0021,  ...,  0.0378, -1.1656, -0.7118],
 ...,
 [-0.3804,  2.4813, -0.3792,  ...,  0.4792, -1.1881,  0.1713],
 [ 0.1178,  2.8472, -0.2366,  ...,  0.4472, -1.2040,  0.5282],
 [ 0.1187,  2.9237, -0.0591,  ...,  0.5884, -1.3009,  0.9568]],
grad_fn=<AddmmBackward0>), view_11: tensor([[[ 0.7109,  2.4857,  1.1
  [ 0.4070,  1.9528, -1.1921,  ...,  0.7105, -0.8874, -0.2245],
  [-1.3988,  2.6029, -2.0021,  ...,  0.0378, -1.1656, -0.7118],
  ...,
  [-0.3804,  2.4813, -0.3792,  ...,  0.4792, -1.1881,  0.1713],
  [ 0.1178,  2.8472, -0.2366,  ...,  0.4472, -1.2040,  0.5282],
  [ 0.1187,  2.9237, -0.0591,  ...,  0.5884, -1.3009,  0.9568]]],
grad_fn=<ViewBackward0>), mul: tensor([[[ 0.3554,  1.2429,  0.5719,
  [ 0.2035,  0.9764, -0.5960,  ...,  0.3552, -0.4437, -0.1123],
  [-0.6994,  1.3014, -1.0010,  ...,  0.0189, -0.5828, -0.3559],
  ...,
  [-0.1902,  1.2406, -0.1896,  ...,  0.2396, -0.5941,  0.0857],
  [ 0.0589,  1.4236, -0.1183,  ...,  0.2236, -0.6020,  0.2641],
  [ 0.0594,  1.4618, -0.0295,  ...,  0.2942, -0.6505,  0.4784]]],
grad_fn=<MulBackward0>), pow_2: tensor([[[ 3.5925e-01,  1.5359e+01,
   -1.1191e+01, -1.1857e-01],
  [ 6.7398e-02,  7.4474e+00, -1.6941e+00,  ...,  3.5865e-01,
   -6.9869e-01, -1.1318e-02],
  [-2.7367e+00,  1.7634e+01, -8.0251e+00,  ...,  5.4199e-05,
   -1.5836e+00, -3.6060e-01],
  ...,
  [-5.5053e-02,  1.5276e+01, -5.4544e-02,  ...,  1.1005e-01,
   -1.6771e+00,  5.0301e-03],
  [ 1.6348e-03,  2.3082e+01, -1.3252e-02,  ...,  8.9408e-02,
   -1.7453e+00,  1.4735e-01],
  [ 1.6729e-03,  2.4992e+01, -2.0605e-04,  ...,  2.0375e-01,
   -2.2017e+00,  8.7586e-01]]], grad_fn=<PowBackward0>), mul_1: tens
   -5.0041e-01, -5.3018e-03],
  [ 3.0137e-03,  3.3301e-01, -7.5751e-02,  ...,  1.6037e-02,
   -3.1242e-02, -5.0607e-04],
  [-1.2237e-01,  7.8850e-01, -3.5884e-01,  ...,  2.4235e-06,
   -7.0813e-02, -1.6124e-02],
  ...,
  [-2.4617e-03,  6.8308e-01, -2.4389e-03,  ...,  4.9210e-03,
   -7.4992e-02,  2.2492e-04],
  [ 7.3101e-05,  1.0321e+00, -5.9256e-04,  ...,  3.9979e-03,
   -7.8042e-02,  6.5886e-03],
  [ 7.4805e-05,  1.1175e+00, -9.2135e-06,  ...,  9.1108e-03,
   -9.8449e-02,  3.9164e-02]]], grad_fn=<MulBackward0>), add_11: ten
  [ 0.4100,  2.2859, -1.2678,  ...,  0.7265, -0.9186, -0.2250],
  [-1.5211,  3.3914, -2.3609,  ...,  0.0378, -1.2364, -0.7279],
  ...,
  [-0.3829,  3.1643, -0.3817,  ...,  0.4841, -1.2631,  0.1716],
  [ 0.1179,  3.8793, -0.2372,  ...,  0.4512, -1.2820,  0.5348],
  [ 0.1188,  4.0412, -0.0591,  ...,  0.5975, -1.3994,  0.9959]]],
grad_fn=<AddBackward0>), mul_2: tensor([[[ 0.5800,  2.5313,  0.9660,
  [ 0.3371,  1.8229, -1.0116,  ...,  0.5787, -0.7339, -0.1785],
  [ 0.3271,  1.8239, -1.0116,  ...,  0.5797, -0.7329, -0.1795],
  [-1.2137,  2.7059, -1.8837,  ...,  0.0302, -0.9865, -0.5808],
  ...,
  [-0.3055,  2.5248, -0.3045,  ...,  0.3863, -1.0078,  0.1369],
  [ 0.0941,  3.0953, -0.1893,  ...,  0.3600, -1.0229,  0.4267],
  [ 0.0948,  3.2244, -0.0471,  ...,  0.4768, -1.1165,  0.7946]]],
grad_fn=<MulBackward0>), tanh: tensor([[[ 0.5227,  0.9874,  0.7469,
  [ 0.3159,  0.9492, -0.7664,  ...,  0.5224, -0.6249, -0.1776],
  [-0.8378,  0.9911, -0.9548,  ...,  0.0302, -0.7559, -0.5232],
  ...,
  [-0.2963,  0.9873, -0.2955,  ...,  0.3682, -0.7649,  0.1360],
  [ 0.0938,  0.9959, -0.1871,  ...,  0.3452, -0.7711,  0.4025],
  [ 0.0945,  0.9968, -0.0471,  ...,  0.4437, -0.8064,  0.6610]]],
grad_fn=<TanhBackward0>), add_12: tensor([[[1.5227, 1.9874, 1.7469,
  [1.3159, 1.9492, 0.2336,  ..., 1.5224, 0.3751, 0.8224],
  [0.1622, 1.9911, 0.0452,  ..., 1.0302, 0.2441, 0.4768],
  ...,
  [0.7037, 1.9873, 0.7045,  ..., 1.3682, 0.2351, 1.1360],
  [1.0938, 1.9959, 0.8129,  ..., 1.3452, 0.2289, 1.4025],
  [1.0945, 1.9968, 0.9529,  ..., 1.4437, 0.1936, 1.6610]]],
grad_fn=<AddBackward0>), mul_3: tensor([[[ 0.5412,  2.4701,  0.9990,
  [ 0.2678,  1.9033, -0.1392,  ...,  0.5408, -0.1664, -0.0923],
  [-0.1135,  2.5913, -0.0452,  ...,  0.0195, -0.1423, -0.1697],
  ...,
  [-0.1338,  2.4654, -0.1336,  ...,  0.3278, -0.1397,  0.0973],
  [ 0.0644,  2.8414, -0.0962,  ...,  0.3008, -0.1378,  0.3704],
  [ 0.0650,  2.9191, -0.0281,  ...,  0.4248, -0.1260,  0.7946]]],
grad_fn=<MulBackward0>), size_16: torch.Size([1, 1024, 10240]), geti
 [ 0.2678,  1.9033, -0.1392,  ...,  0.5408, -0.1664, -0.0923],
 [-0.1135,  2.5913, -0.0452,  ...,  0.0195, -0.1423, -0.1697],
 ...,
 [-0.1338,  2.4654, -0.1336,  ...,  0.3278, -0.1397,  0.0973],
 [ 0.0644,  2.8414, -0.0962,  ...,  0.3008, -0.1378,  0.3704],
 [ 0.0650,  2.9191, -0.0281,  ...,  0.4248, -0.1260,  0.7946]],
grad_fn=<ViewBackward0>), transformer_h_0_mlp_c_proj_weight: tensor(
 [ 0.0168,  0.0085, -0.0274,  ..., -0.0099,  0.0177,  0.0222],
 [-0.0040, -0.0141, -0.0008,  ..., -0.0056,  0.0105,  0.0092],
 ...,
 [-0.0171, -0.0042, -0.0091,  ..., -0.0148, -0.0021, -0.0234],
 [-0.0006, -0.0063,  0.0287,  ..., -0.0038,  0.0116, -0.0057],
 [ 0.0150,  0.0035,  0.0035,  ...,  0.0155,  0.0075, -0.0009]]), add
 [ 0.4927,  0.6437,  1.6678,  ...,  0.5361,  0.9436, -1.2700],
 [ 0.0709,  1.2535,  1.8534,  ..., -0.2436,  0.3765, -0.0302],
 ...,
 [ 1.3752,  2.2154,  2.3412,  ...,  1.0088,  0.9718,  0.0143],
 [ 1.4167,  1.8492,  2.4114,  ...,  0.2166,  0.5827,  0.4021],
 [ 1.5392,  1.8969,  1.5351,  ...,  0.7653,  0.3408,  0.5918]],
grad_fn=<AddmmBackward0>), view_13: tensor([[[-0.6075,  0.0314,  2.8
  [ 0.4927,  0.6437,  1.6678,  ...,  0.5361,  0.9436, -1.2700],
  [ 0.0709,  1.2535,  1.8534,  ..., -0.2436,  0.3765, -0.0302],
  ...,
  [ 1.3752,  2.2154,  2.3412,  ...,  1.0088,  0.9718,  0.0143],
  [ 1.4167,  1.8492,  2.4114,  ...,  0.2166,  0.5827,  0.4021],
  [ 1.5392,  1.8969,  1.5351,  ...,  0.7653,  0.3408,  0.5918]]],
grad_fn=<ViewBackward0>), transformer_h_0_mlp_dropout: tensor([[[-0.
  [ 0.5474,  0.7153,  0.0000,  ...,  0.5956,  1.0485, -1.4111],
  [ 0.0788,  1.3928,  2.0593,  ..., -0.2707,  0.4183, -0.0335],
  ...,
  [ 1.5280,  2.4616,  2.6013,  ...,  1.1209,  1.0797,  0.0159]
  [ 1.5280,  2.4616,  2.6013,  ...,  1.1209,  1.0797,  0.0159],
  [ 1.5741,  2.0547,  2.6793,  ...,  0.2406,  0.6475,  0.4468],
  [ 1.7102,  2.1077,  1.7057,  ...,  0.8504,  0.3787,  0.6576]]],
grad_fn=<MulBackward0>), add_14: tensor([[[-0.0676,  0.8354,  0.9852
  [ 0.7776,  1.3505,  0.0131,  ...,  1.0794,  1.7959, -1.9005],
  [ 0.7805,  1.8984,  1.4470,  ..., -0.2796,  1.5880, -0.0335],
  ...,
  [ 2.0118,  2.6637,  1.8979,  ...,  0.6701,  1.9714, -0.4240],
  [ 2.0542,  2.2219,  1.8859,  ..., -0.2415,  1.5152,  0.0469],
  [ 2.2411,  2.0469,  0.9297,  ...,  0.4799,  0.3735,  0.2159]]],
grad_fn=<AddBackward0>), transformer_ln_f: tensor([[[-0.0672,  0.644
  [ 0.6613,  1.1444,  0.0168,  ...,  0.9158,  1.5198, -1.5966],
  [ 0.6541,  1.6069,  1.2222,  ..., -0.2494,  1.3423, -0.0397],
  ...,
  [ 1.7757,  2.3598,  1.6737,  ...,  0.5736,  1.7395, -0.4065],
  [ 1.8481,  1.9996,  1.6962,  ..., -0.2257,  1.3612,  0.0349],
  [ 2.0200,  1.8435,  0.8287,  ...,  0.4200,  0.3234,  0.1803]]],
grad_fn=<NativeLayerNormBackward0>), view_14: tensor([[[-0.0672,  0.
  [ 0.6613,  1.1444,  0.0168,  ...,  0.9158,  1.5198, -1.5966],
  [ 0.6541,  1.6069,  1.2222,  ..., -0.2494,  1.3423, -0.0397],
  ...,
  [ 1.7757,  2.3598,  1.6737,  ...,  0.5736,  1.7395, -0.4065],
  [ 1.8481,  1.9996,  1.6962,  ..., -0.2257,  1.3612,  0.0349],
  [ 2.0200,  1.8435,  0.8287,  ...,  0.4200,  0.3234,  0.1803]]],
grad_fn=<ViewBackward0>), lm_head: tensor([[[ 0.1186,  1.5185,  0.35
  [-0.9143,  0.6947,  1.0530,  ..., -0.4587, -0.3837,  0.1027],
  [-0.5018,  0.5240,  1.2935,  ...,  0.5907,  0.8032, -1.3382],
  ...,
  [-0.0990, -0.2428,  1.3203,  ..., -1.9470, -0.6983, -0.3977],
  [-0.1986,  0.3236,  1.6286,  ..., -0.6152, -0.5064, -0.0880],
  [-0.0306,  1.4860,  1.9004,  ..., -1.3813, -0.6709,  0.2560]]],
grad_fn=<UnsafeViewBackward0>), output: ({'logits': tensor([[[ 0.118
  [-0.9143,  0.6947,  1.0530,  ..., -0.4587, -0.3837,  0.1027],
  [-0.5018,  0.5240,  1.2935,  ...,  0.5907,  0.8032, -1.3382],
  ...,
  [-0.0990, -0.2428,  1.3203,  ..., -1.9470, -0.6983, -0.3977],
  [-0.1986,  0.3236,  1.6286,  ..., -0.6152, -0.5064, -0.0880],
  [-0.0306,  1.4860,  1.9004,  ..., -1.3813, -0.6709,  0.2560]]],
grad_fn=<UnsafeViewBackward0>), 'past_key_values': ((tensor([[[[-0.0
   [-0.2898, -0.9153,  0.2059,  ..., -0.7287, -1.3031,  0.2322],
   [ 0.8982,  0.5432,  0.0756,  ...,  1.0897, -1.7253,  0.2846],
   ...,
   [ 0.7235, -0.1132, -0.0178,  ..., -0.8767,  0.9113,  1.0387],
   [ 0.4877,  0.9024,  0.1619,  ..., -0.7393, -1.1445, -0.2149],
   [-0.3400, -0.1774,  1.1426,  ..., -1.6697,  0.7312,  0.2506]],

  [[-1.1550, -0.6857, -0.0548,  ...,  0.7911, -1.2452,  1.1871],
   [-1.4128,  0.3902,  0.5208,  ...,  1.5560, -1.4400,  0.8671],
   [-0.8195, -1.1852,  0.7765,  ...,  1.4433, -0.8974,  1.2275],
   ...,
   [-1.3198, -0.0586,  0.6382,  ...,  0.1729, -0.2085,  0.6523],
   [-0.1504, -0.2212, -0.6004,  ...,  0.8414,  1.0264,  0.3428],
   [ 0.4139,  0.7714, -0.0446,  ...,  1.4960, -1.7593,  0.9587]],

  [[ 0.9380, -0.1100, -1.4938,  ...,  0.7169,  0.9322,  0.0979],
   [-0.2036,  0.7863,  0.6648,  ...,  0.5309, -0.1505,  0.1302],
   [ 0.5280, -1.0059,  1.2404,  ...,  2.1611,  0.4663, -2.0759],
   ...,
   [ 0.3867,  1.8113,  0.2210,  ...,  0.1097,  0.8334,  2.3712],
   [ 0.3867,  1.8113, -0.2210,  ..., -0.1997,  0.8334, -2.3713],
   [-0.2839,  1.5350, -0.5067,  ...,  0.4349,  0.4209,  0.6408],
   [-0.2052, -1.3002, -0.2397,  ...,  0.9969,  1.4980, -1.2450]],

  ...,

  [[ 1.5353, -0.6293,  0.1201,  ...,  1.1989,  0.0139,  0.5774],
   [-0.1049, -1.6774,  0.6332,  ...,  1.0535, -0.5428, -0.9433],
   [ 1.0090, -0.6773,  0.5736,  ...,  1.5946,  0.2422, -0.2608],
   ...,
   [ 0.9555, -0.3896,  1.3091,  ...,  1.7802,  0.3060,  0.6750],
   [ 1.4842, -0.3871,  1.0065,  ...,  2.0689, -1.3078, -0.7658],
   [-0.1303, -2.1165,  1.7418,  ...,  1.6911,  0.4163, -1.4566]],

  [[-1.7977, -0.0972,  0.2891,  ..., -0.4595, -1.0992,  0.6587],
   [-0.7844,  0.0462, -1.1205,  ...,  0.4437, -0.6814,  1.1193],
   [-0.6200,  0.4880, -1.5203,  ...,  0.4020, -0.7304, -0.3302],
   ...,
   [-1.2726, -0.3915,  0.6958,  ...,  0.4698, -1.2266, -0.3855],
   [-0.9818,  0.4404, -0.4635,  ...,  0.2169, -1.1535,  0.0350],
   [-0.8389, -0.8111, -0.7810,  ..., -0.4379, -0.9150,  1.7493]],

  [[ 0.8641,  0.5925,  0.0250,  ...,  0.9172,  0.8187,  0.0937],
   [ 1.1099,  1.4441,  0.7351,  ...,  1.5720, -0.0148,  0.3059],
   [ 0.5323,  0.8540,  1.2001,  ...,  0.8550,  0.7815, -0.1665],
   ...,
   [ 0.1825, -0.2630, -0.8271,  ...,  0.5835,  0.7090, -1.3823],
   [-0.5885, -1.2048,  0.7633,  ...,  2.2055,  2.1243,  1.1146],
   [ 1.2332, -0.3148,  1.6390,  ...,  1.5862,  1.8645, -0.2542]]]],
grad_fn=<PermuteBackward0>), tensor([[[[ 5.4475e-01,  4.1136e-01,  6
     5.3991e-02, -1.1233e+00],
   [ 1.4472e-01,  1.5402e-01, -4.0394e-01,  ...,  5.9276e-01,
    -1.3165e-01, -1.2211e+00],
   [ 1.3474e+00, -1.0317e-02, -6.0596e-01,  ..., -5.1366e-01,
    -1.3814e+00, -6.5122e-01],
   ...,
   [ 3.4586e-01, -1.0730e+00,  9.3181e-01,  ...,  5.9873e-01,
     2.6745e-01, -6.3752e-01],
   [-1.0901e+00, -1.2123e+00, -5.5091e-01,  ..., -1.1037e+00,
    -1.9915e+00,  1.9754e-01],
   [ 1.0175e+00, -3.5308e-01,  7.2218e-02,  ...,  1.3013e+00,
    -2.2913e-01, -4.9519e-01]],

  [[-3.1268e-01,  7.5064e-01,  1.3438e+00,  ..., -1.2461e-01,
     4.1225e-01,  1.0110e+00],
   [ 1.0750e+00, -1.0887e+00,  6.9819e-01,  ...,  6.7687e-01,
     1.4665e-01,  1.2630e+00],
   [ 8.0351e-01,  9.6109e-01,  1.4139e+00,  ...,  1.4535e+00,
     9.8968e-01,  4.2794e-01],
   ...,
   [ 1.9807e-01,  8.4666e-01,  1.7526e+00,  ..., -5.1485e-01,
     1.5930e+00, -1.3223e-01],
   [-1.3602e-01,  6.1435e-01,  1.5291e+00,  ...,  1.2577e+00,
     8.6335e-01,  1.5900e+00],
   [-1.6743e-01,  5.3940e-01,  1.3644e+00,  ..., -4.3406e-01,
     2.0700e-02, -1.2070e-01]],

  [[ 5.3532e-01, -1.7978e-01, -9.8737e-01,  ...,  7.4715e-01,
     4.8242e-01,  1.5642e+00]
    -4.8342e-01,  1.5643e+00],
   [ 8.3065e-01, -1.6116e-01,  5.3616e-02,  ...,  1.6874e+00,
    -9.2703e-01, -4.5544e-01],
   [ 1.4144e-01, -6.9275e-01, -1.0767e+00,  ...,  2.9450e-01,
    -4.7638e-01,  1.6170e+00],
   ...,
   [-2.0230e-01, -2.1122e+00, -5.2980e-01,  ...,  2.0098e+00,
    -1.2939e+00,  1.3258e+00],
   [ 9.6869e-01, -7.9333e-01, -2.3843e-01,  ...,  3.0226e-01,
    -6.3827e-03,  8.3277e-01],
   [ 4.9090e-02, -1.1509e+00, -8.9348e-01,  ..., -6.8726e-01,
    -6.2417e-01, -2.0275e-01]],

  ...,

  [[-6.7070e-01, -9.6281e-01,  3.8247e-01,  ...,  1.4481e+00,
    -6.7513e-01,  7.8456e-01],
   [-2.3104e-01, -1.3227e+00, -9.4377e-01,  ...,  6.0654e-01,
    -9.6277e-01,  1.0423e+00],
   [ 1.8056e-01, -3.9629e-01, -1.1429e-03,  ...,  7.0036e-01,
    -8.7710e-01,  6.5044e-01],
   ...,
   [ 5.5298e-02, -1.6574e+00,  1.5176e+00,  ...,  5.3913e-01,
    -8.6454e-02, -7.3583e-01],
   [ 1.4602e+00, -1.6949e+00,  5.7760e-01,  ...,  2.7545e+00,
     1.1102e-01,  3.2058e-01],
   [ 2.2913e-01, -1.8145e+00,  6.4851e-01,  ...,  2.1424e+00,
    -2.9331e-01,  5.9818e-01]],

  [[ 3.2699e-02, -1.7002e+00,  9.3860e-01,  ...,  1.0049e+00,
     1.3244e+00, -1.6203e+00],
   [ 5.4551e-01,  3.4231e-01,  3.7528e-01,  ...,  1.2852e+00,
     1.0643e+00, -1.3481e-03],
   [ 1.2079e+00, -1.7680e+00,  1.5530e+00,  ...,  1.8241e+00,
     3.9575e-01, -5.3557e-02],
   ...,
   [ 2.0102e-01, -1.4970e-01,  3.4424e-02,  ...,  6.0766e-01,
     9.4239e-01,  2.3055e-01],
   [-3.7983e-02, -1.2483e-01,  1.6051e+00,  ...,  3.6712e-01,
     1.5149e+00, -2.0265e-01],
   [ 1.5226e+00, -1.0960e+00,  1.1153e+00,  ...,  1.7604e+00,
     5.2075e-01,  2.7665e-02]],

  [[-2.1535e-01, -2.6749e+00,  2.0272e-01,  ..., -2.6837e-01,
     1.2549e-01,  9.6314e-01],
   [-4.1424e-01, -1.1618e+00,  3.8963e-02,  ..., -1.5521e-04,
    -1.2065e-01,  3.4900e-01],
   [ 2.9455e-01, -2.7297e+00,  4.7883e-01,  ...,  3.2143e-01,
     1.0082e+00, -6.9631e-01],
   ...,
   [ 3.8988e-01, -6.9052e-01,  1.0939e+00,  ..., -5.6343e-01,
    -8.9841e-01,  8.3196e-01],
   [ 6.6216e-01, -9.5437e-01, -5.0103e-01,  ...,  5.7716e-01,
     7.4068e-01,  1.7688e-01],
   [-1.7653e+00, -2.9693e+00,  1.2386e+00,  ...,  7.7132e-01,
     2.3493e-01,  1.0911e+00]]]], grad_fn=<PermuteBackward0>)),)},)}
```




size (0.00 ms)
[1, 1024]
getitem (0.00 ms)
getitem_2 (0.00 ms)
getitem_3 (0.00 ms)
view (0.03 ms)
[1, 1024]
[1, 1024]
[1, 1024]
add (0.00 ms)
size_1 (0.00 ms)
getattr_1 (0.00 ms)
transformer_wte (2.67 ms)
getitem_1 (0.00 ms)
arange (0.00 ms)
[1024]
unsqueeze (0.00 ms)
[1, 1024, 2560]
[1, 1024]
view_1 (0.00 ms)
[1, 1024]
transformer_wpe (3.66 ms)
[1, 1024, 2560]
add_1 (5.21 ms)
[1, 1024, 2560]
transformer_drop (11.61 ms)
[1, 1024, 2560]
[1, 1024, 2560]
size_2 (0.00 ms)
transformer_h_0_ln_1 (4.36 ms)
[1, 1024, 2560]
[1, 1024, 2560]
[1, 1024, 2560]
add_2 (0.00 ms)
size_3 (0.00 ms)
size_4 (0.00 ms)
getitem_4 (0.00 ms)
transformer_h_0_attn_c_attn_bias (0.00 ms)
transformer_h_0_attn_c_attn_weight (0.00 ms)
view_2 (0.00 ms)
[7680]
[2560, 7680]
[1024, 2560]
add_3 (0.00 ms)
addmm (546.72 ms)
[1024, 7680]
view_3 (0.00 ms)
[1, 1024, 7680]
split (0.01 ms)
getitem_6 (0.00 ms)
getitem_5 (0.00 ms)
[1, 1024, 2560]
[1, 1024, 2560]
size_5 (0.00 ms)
getitem_7 (0.00 ms)
size_6 (0.00 ms)
[1, 1024, 2560]
getitem_9 (0.00 ms)
[1, 1024, 2560]
[1, 1024, 2560]
getitem_8 (0.00 ms)
size_7 (0.00 ms)
add_5 (0.00 ms)
add_4 (0.00 ms)
getitem_10 (0.00 ms)
[1, 1024, 2560]
view_5 (0.00 ms)
view_4 (0.00 ms)
add_6 (0.00 ms)
[1, 1024, 32, 80]
[1, 1024, 32, 80]
permute_1 (0.00 ms)
permute (0.00 ms)
view_6 (0.00 ms)
[1, 32, 1024, 80]
[1, 32, 1024, 80]
[1, 32, 1024, 80]
[1, 1024, 32, 80]
size_10 (0.00 ms)
transpose (0.00 ms)
size_9 (0.00 ms)
[1, 32, 1024, 80]
permute_2 (0.00 ms)
[1, 32, 80, 1024]
[1, 32, 1024, 80]
[1, 32, 1024, 80]
sub (0.00 ms)
transformer_h_0_attn_bias (0.00 ms)
matmul (211.50 ms)
size_8 (0.00 ms)
[1, 1, 2048, 2048]
[1, 32, 1024, 1024]
[1, 32, 1024, 1024]
getattr_8 (0.00 ms)
getitem_11 (0.01 ms)
getattr_3 (0.00 ms)
getattr_2 (0.00 ms)
pow_1 (0.00 ms)
[1, 1, 1024, 1024]
[1, 32, 1024, 1024]
full (0.01 ms)
to (0.20 ms)
[]


truediv (104.94 ms)
[1, 32, 1024, 1024]
[1, 32, 1024, 1024]
getattr_4 (0.00 ms)
[1, 32, 1024, 1024]
getattr_7 (0.00 ms)
finfo (0.00 ms)
[1, 1024, 2560]
[1, 1, 1024, 1024]
getattr_6 (0.00 ms)
[1, 32, 1024, 1024]
getattr_5 (0.00 ms)
[1, 32, 1024, 80]
full_1 (0.00 ms)
[]
to_1 (0.00 ms)
[]
where (118.56 ms)
[1, 32, 1024, 1024]
softmax (134.53 ms)
[1, 32, 1024, 1024]
type_1 (0.00 ms)
[1, 32, 1024, 1024]
transformer_h_0_attn_attn_dropout (324.40 ms)
[1, 32, 1024, 1024]
matmul_1 (97.45 ms)
[1, 32, 1024, 80]
permute_3 (0.00 ms)
[1, 1024, 32, 80]
contiguous (1.84 ms)
[1, 1024, 32, 80]
size_11 (0.00 ms)
getitem_12 (0.00 ms)
[1, 1024, 32, 80]
add_7 (0.00 ms)
view_7 (0.00 ms)
[1, 1024, 2560]
[1, 1024, 2560]
[1, 1024, 2560]
size_13 (0.00 ms)
size_12 (0.00 ms)
view_8 (0.00 ms)
transformer_h_0_attn_c_proj_bias (0.00 ms)
transformer_h_0_attn_c_proj_weight (0.00 ms)
getitem_13 (0.00 ms)
[1024, 2560]
[2560]
[2560, 2560]
addmm_1 (192.08 ms)
add_8 (0.00 ms)
[1, 32, 1024, 80]
[1024, 2560]
view_9 (0.00 ms)
[1, 32, 1024, 80]
[1, 1024, 2560]
transformer_h_0_attn_resid_dropout (18.17 ms)
[1, 1024, 2560]
add_9 (3.44 ms)
[1, 1024, 2560]
transformer_h_0_ln_2 (4.94 ms)
[1, 1024, 2560]
[1, 1024, 2560]
[1, 1024, 2560]
size_14 (0.00 ms)
size_15 (0.00 ms)
getitem_14 (0.00 ms)
view_10 (0.01 ms)
transformer_h_0_mlp_c_fc_bias (0.00 ms)
transformer_h_0_mlp_c_fc_weight (0.00 ms)
[1024, 2560]
[10240]
[2560, 10240]
add_10 (0.00 ms)
addmm_2 (785.75 ms)



	op	target	args	kwargs	shape	l
0	placeholder	input_ids	()	{}	[1, 1024]	C
1	call_method	size	(input_ids,)	{}	None	C
2	call_function	<built-in function getitem>	(size, -1)	{}	None	C
3	call_method	view	(input_ids, -1, getitem)	{}	[1, 1024]	C
4	call_method	size	(view,)	{}	None	C
...	...	...	...	...	...	
		<built-in	(add_9		[1,	

120	call_function	<built-in function add>	(add_9, transformer_h_0_mlp_dropout)	{}	1024, 2560]	2
121	call_module	transformer.ln_f	(add_14,)	{}	[1, 1024, 2560]	5
122	call_method	view	(transformer_ln_f, add_2)	{}	[1, 1024, 2560]	0
123	call_module	lm_head	(view_14,)	{}	[1, 1024, 50257]	4091
124	output	output	({'logits': lm_head, 'past_key_values': ((perm…	{}	None	0

125 rows × 6 columns

A.3 Graph Analysis II [1 point]

Task

Complete the `findHeavyOps` function in `analysis.py` to identify the most time-consuming operations.

- The function should return the **top 3 nodes** with the highest latency.
- The result should be a **list of tuples** in the format:

```
[(NODE_NAME, LATENCY), ...]
```

```
from analysis import findHeavyOps

print(findHeavyOps(graphmodule))
```

```
[('lm_head', 4091.5608644485474), ('addmm_2', 785.748553276062), ('addmm_3'
```