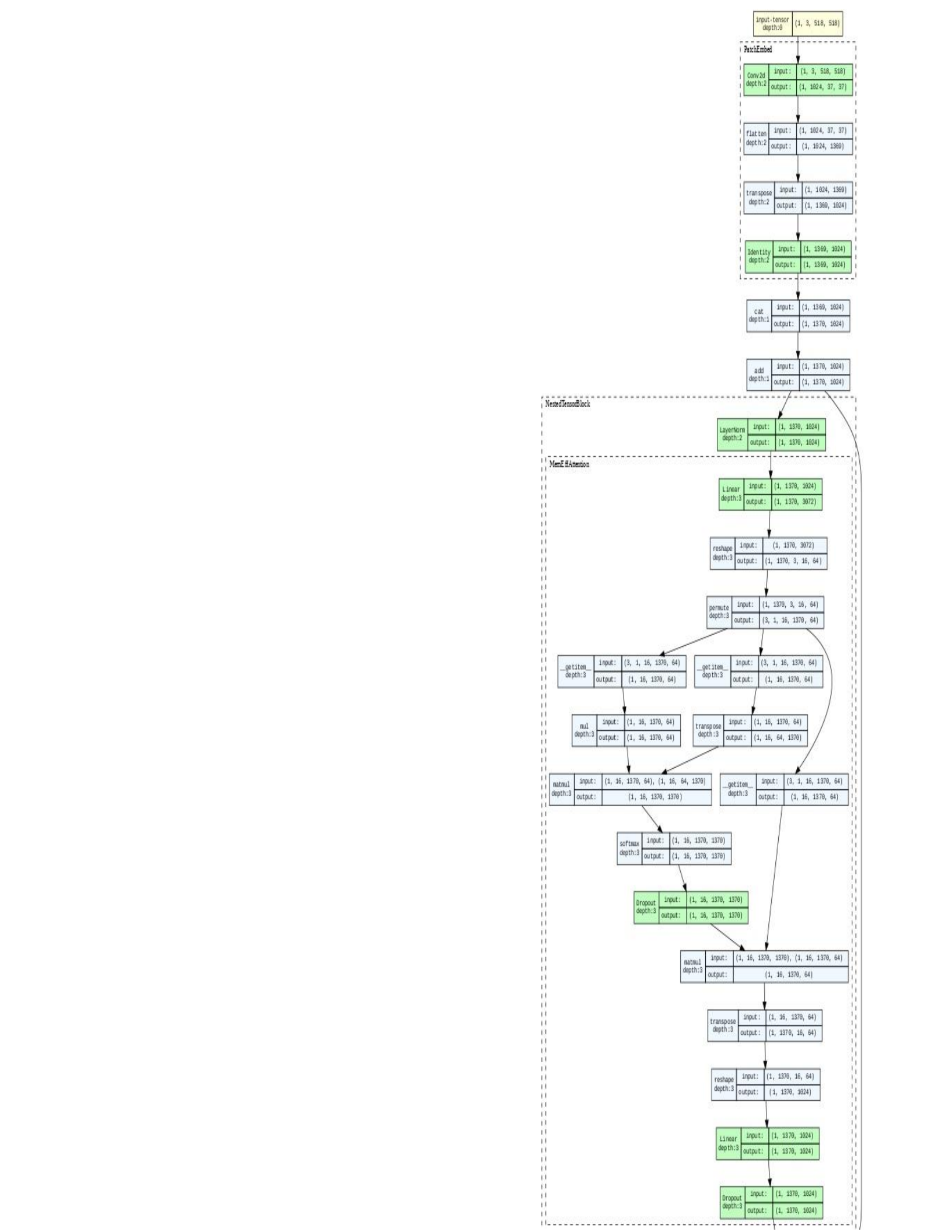
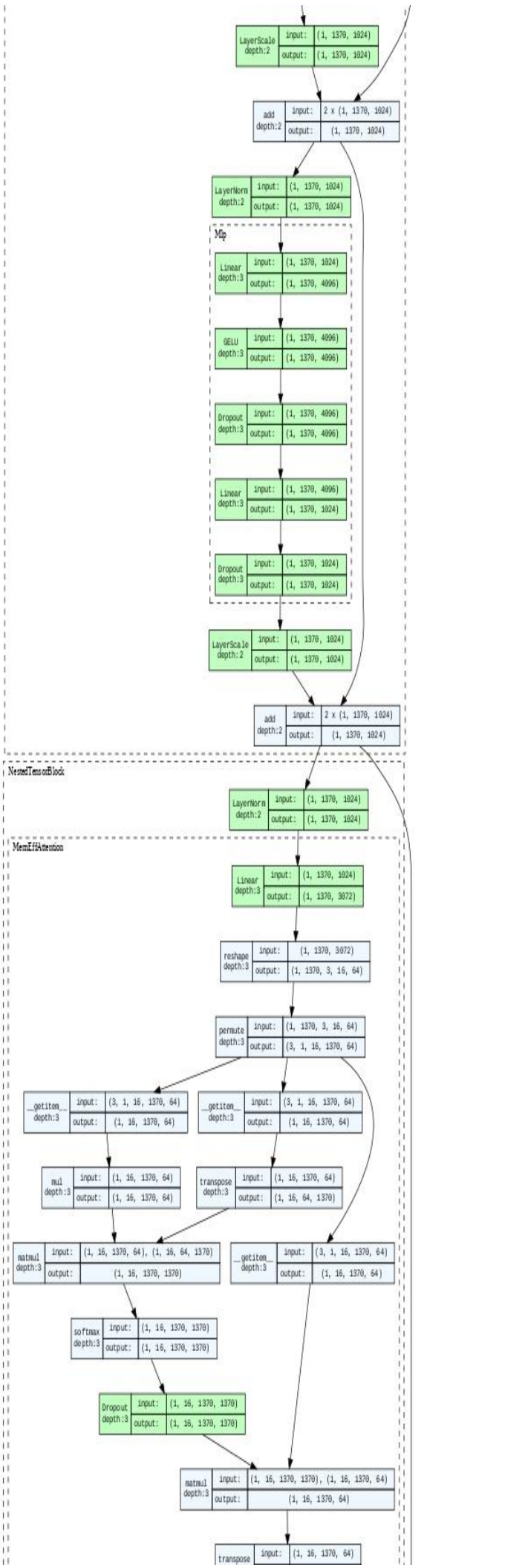
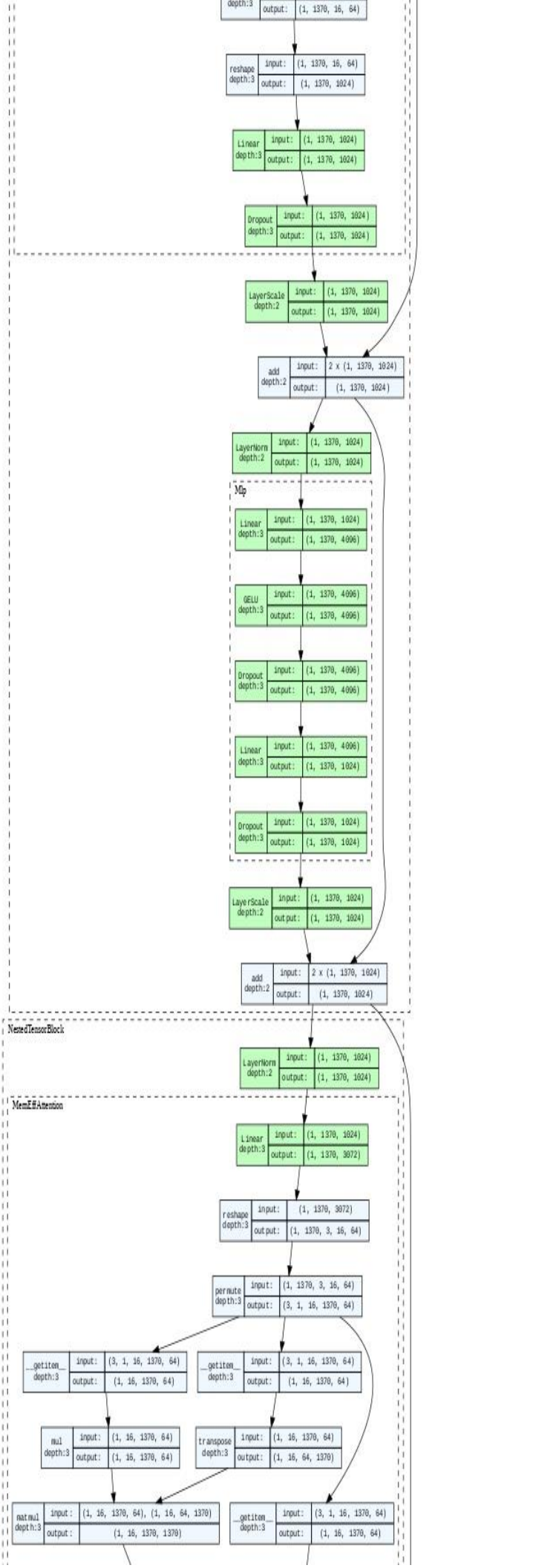
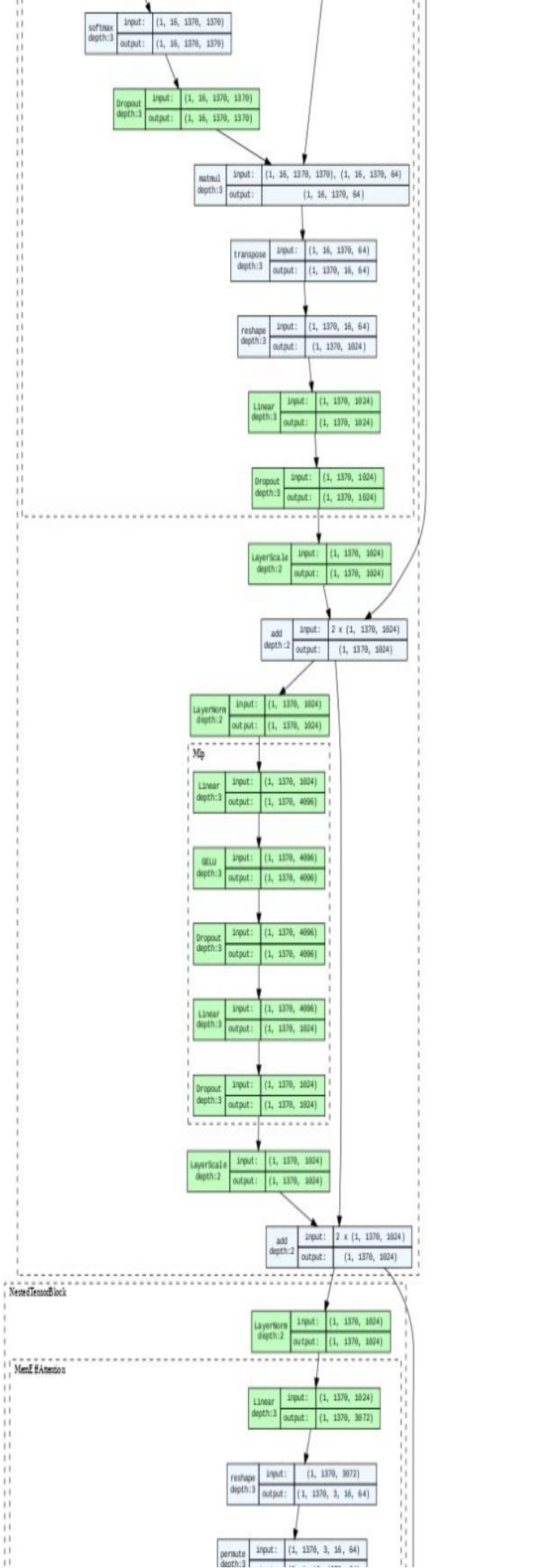
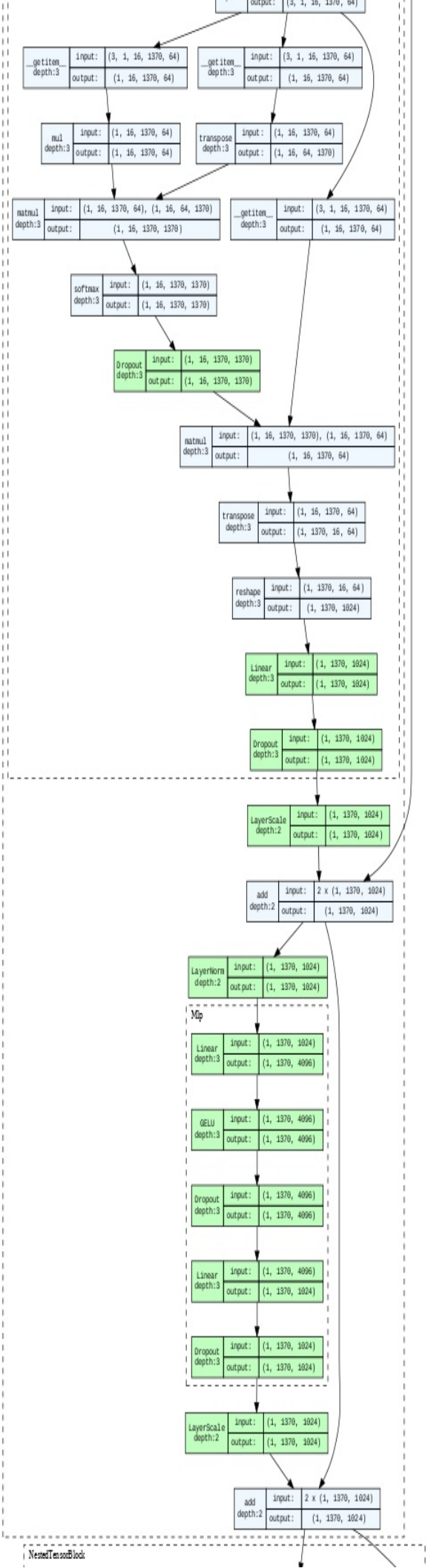
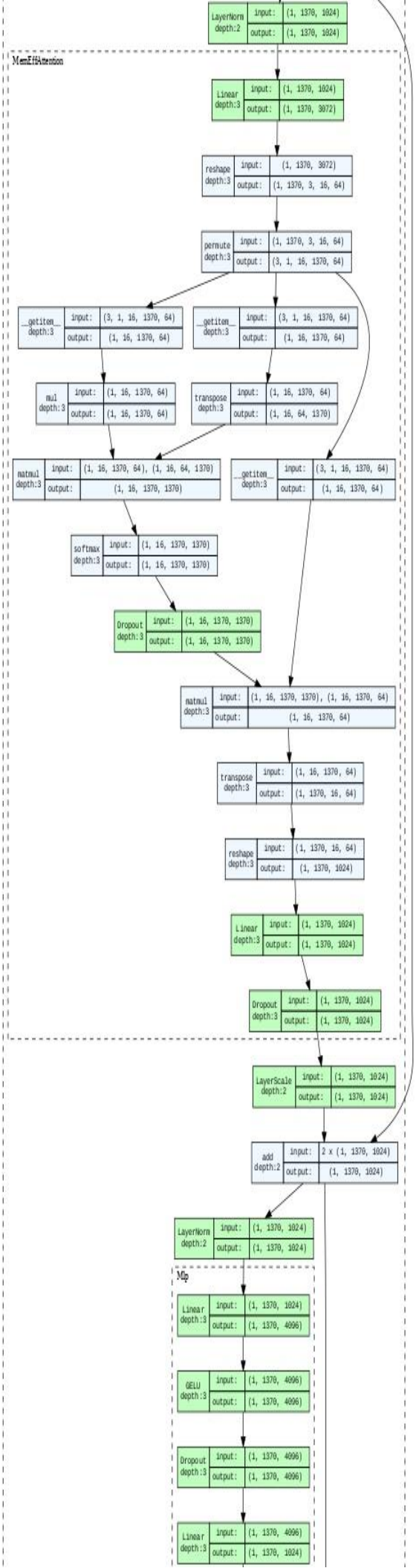

input-tensor depth:0 (1, 3, 518, 518)
PatchEmbed
Conv2d depth:2 input: (1, 3, 518, 518) output: (1, 1024, 37, 37)
flatten depth:2 input: (1, 1024, 37, 37) output: (1, 1024, 1369)
transpose depth:2 input: (1, 1024, 1369) output: (1, 1369, 1024)
Identity depth:2 input: (1, 1369, 1024) output: (1, 1369, 1024)
cat depth:1 input: (1, 1369, 1024) output: (1, 1370, 1024)
add depth:1 input: (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)

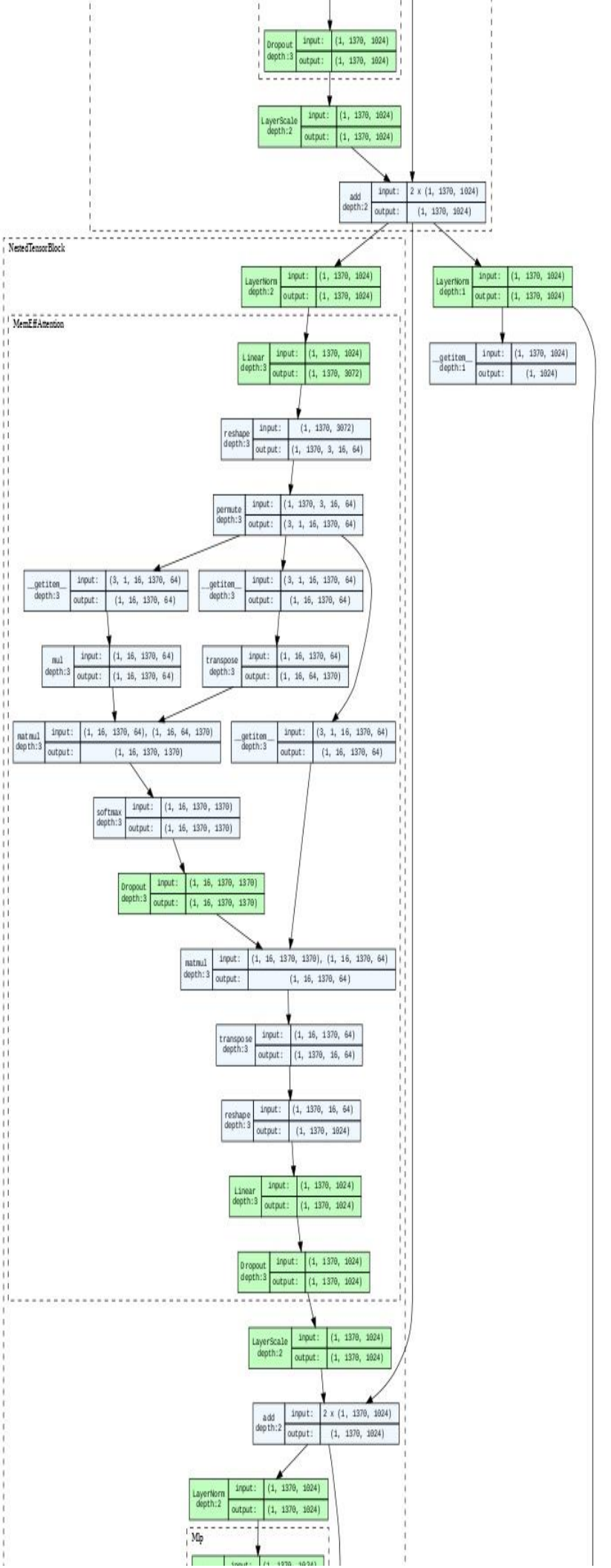

LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose input: (1, 16, 1370, 64)

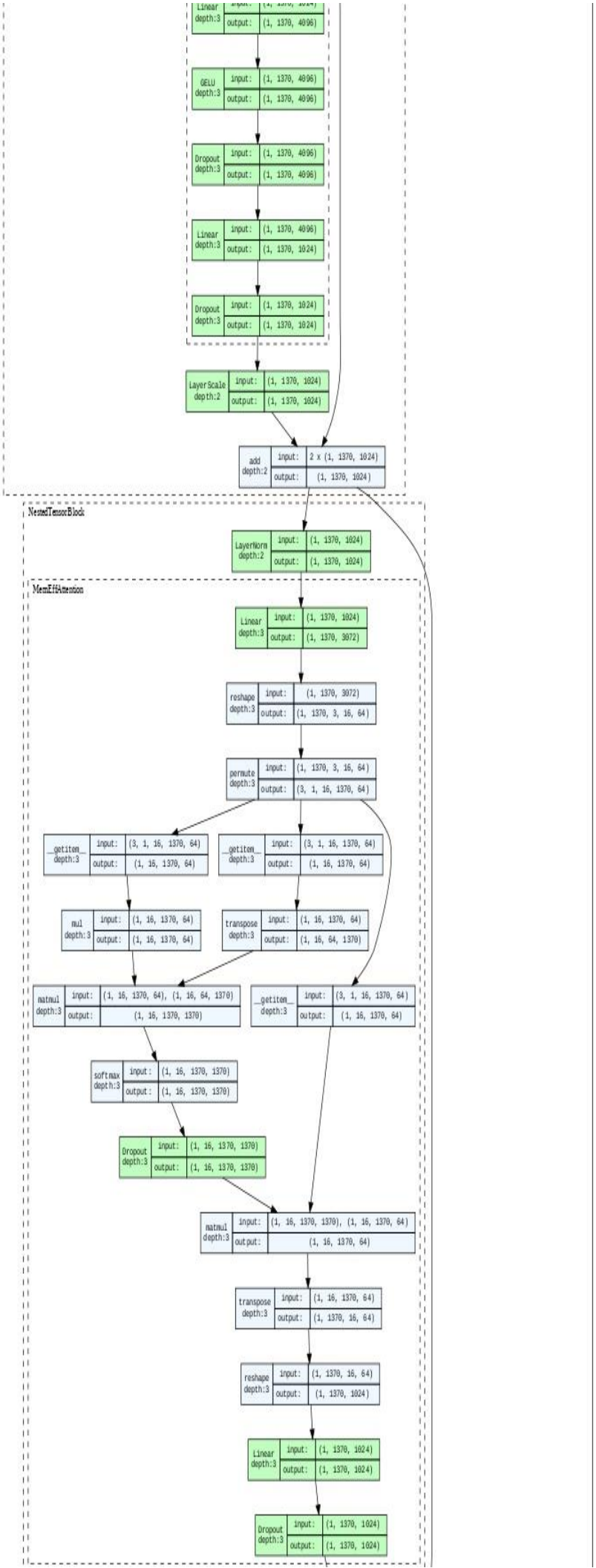

reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)

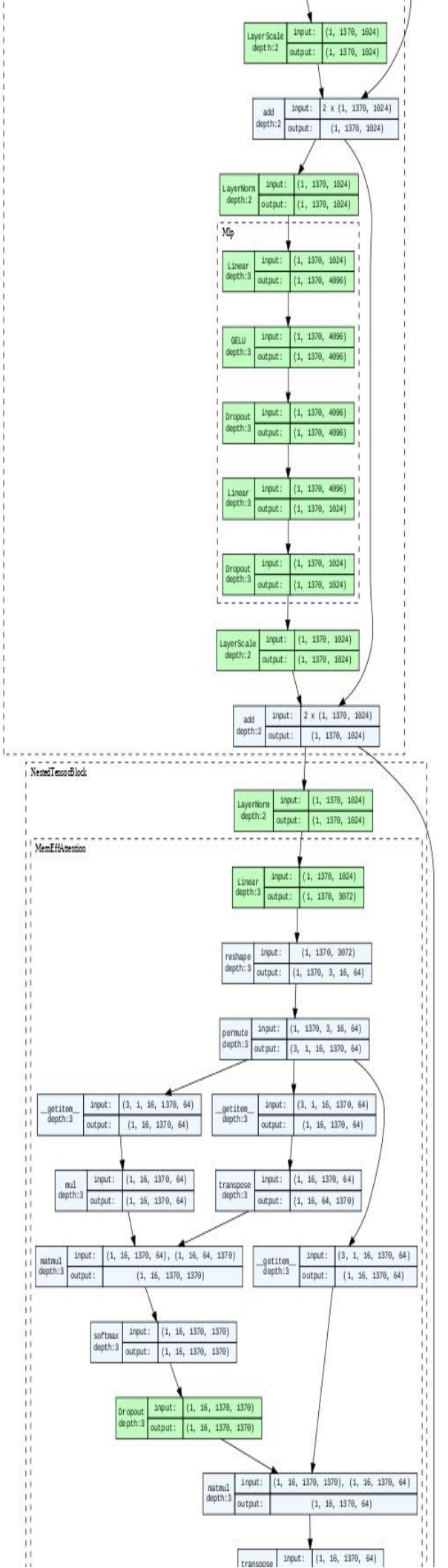

softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370); (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64)

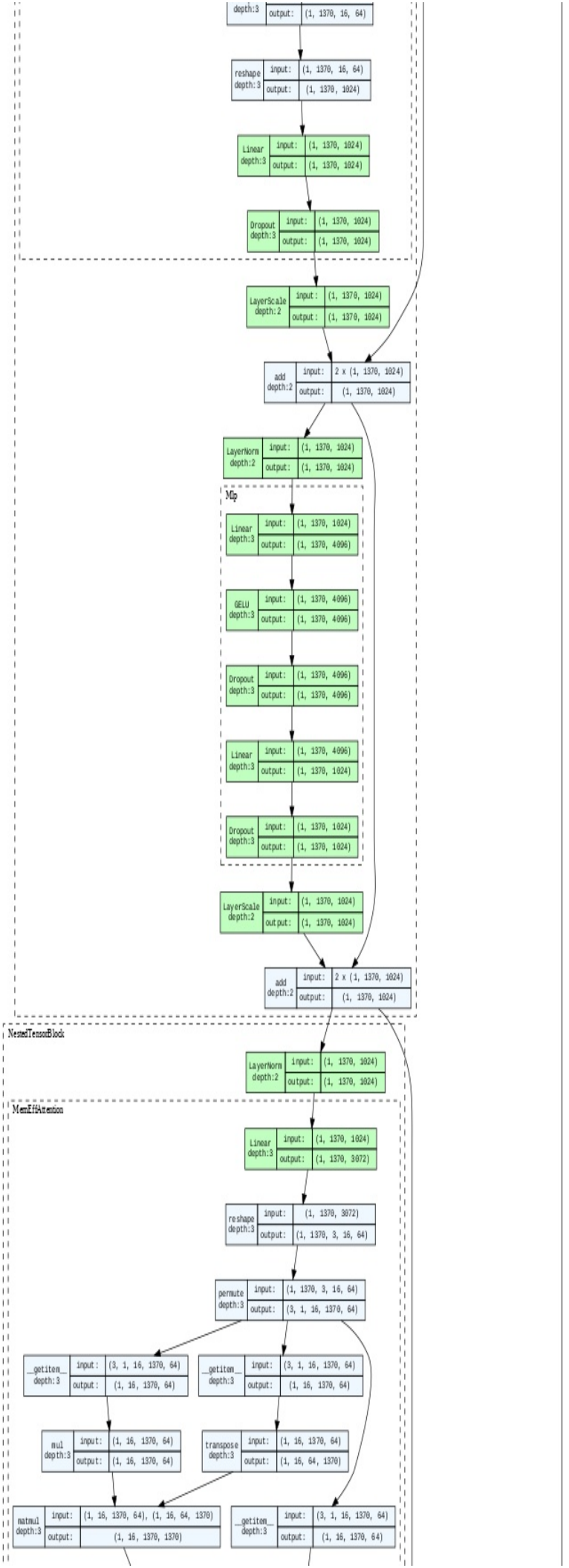

__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock

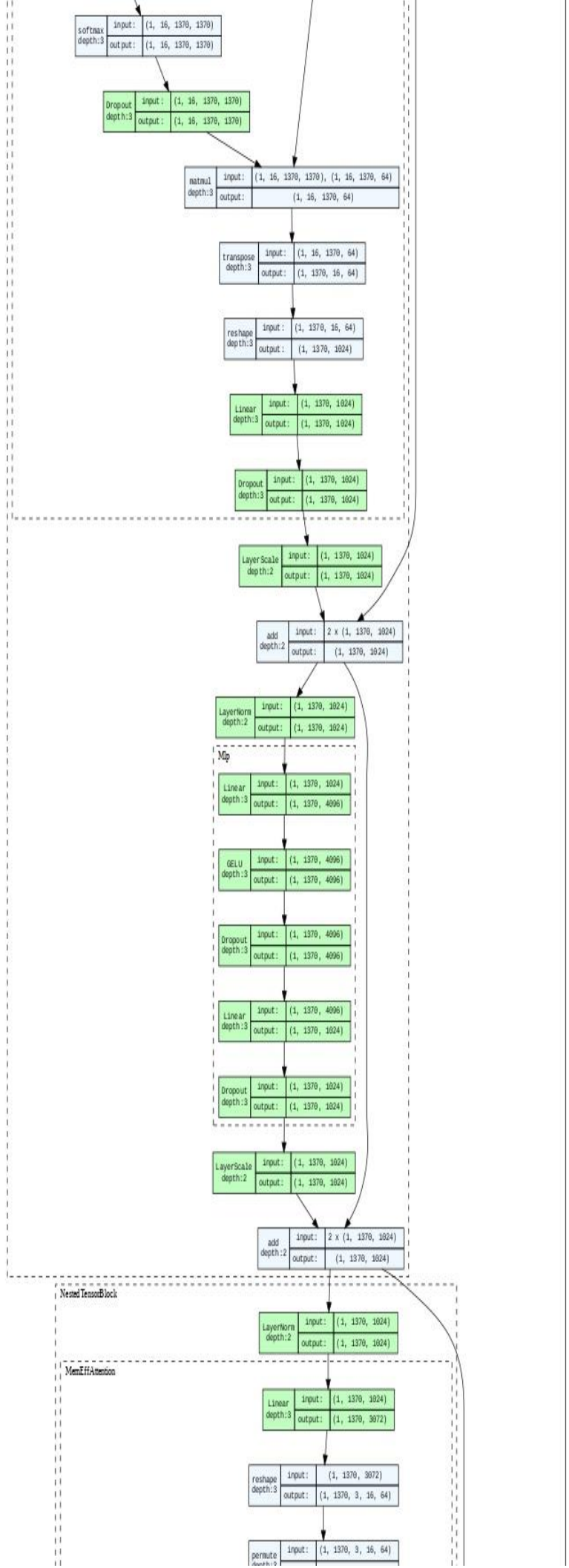

LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)

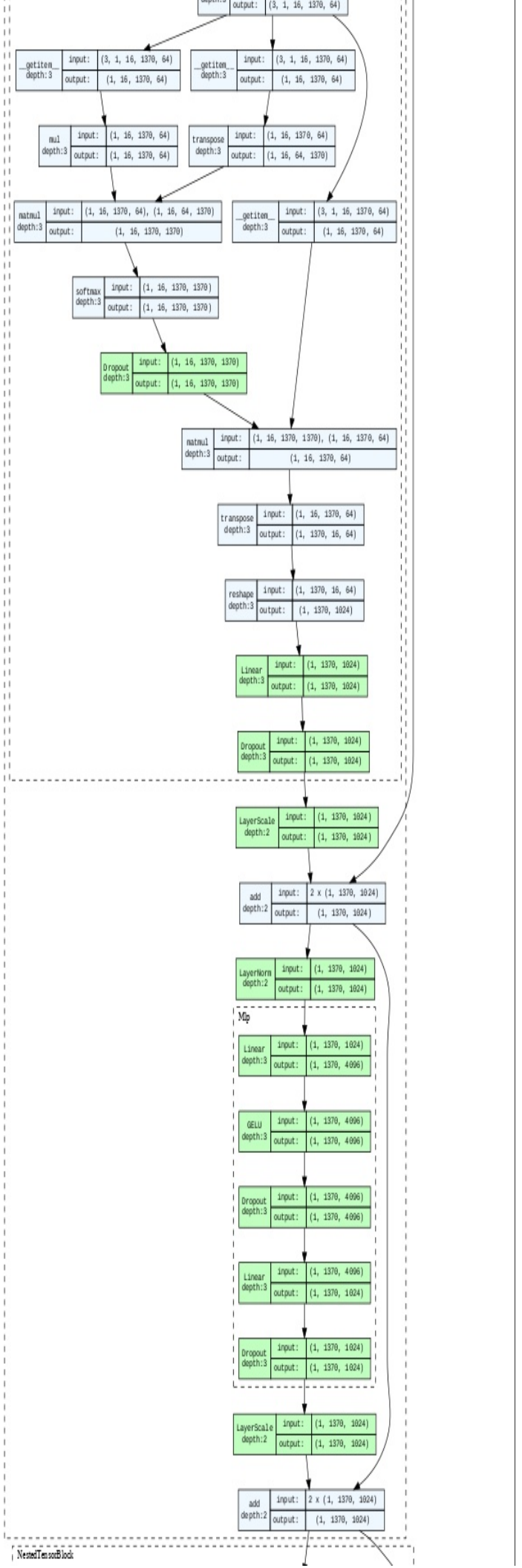

Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:1 input: (1, 1370, 1024) output: (1, 1370, 1024)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp

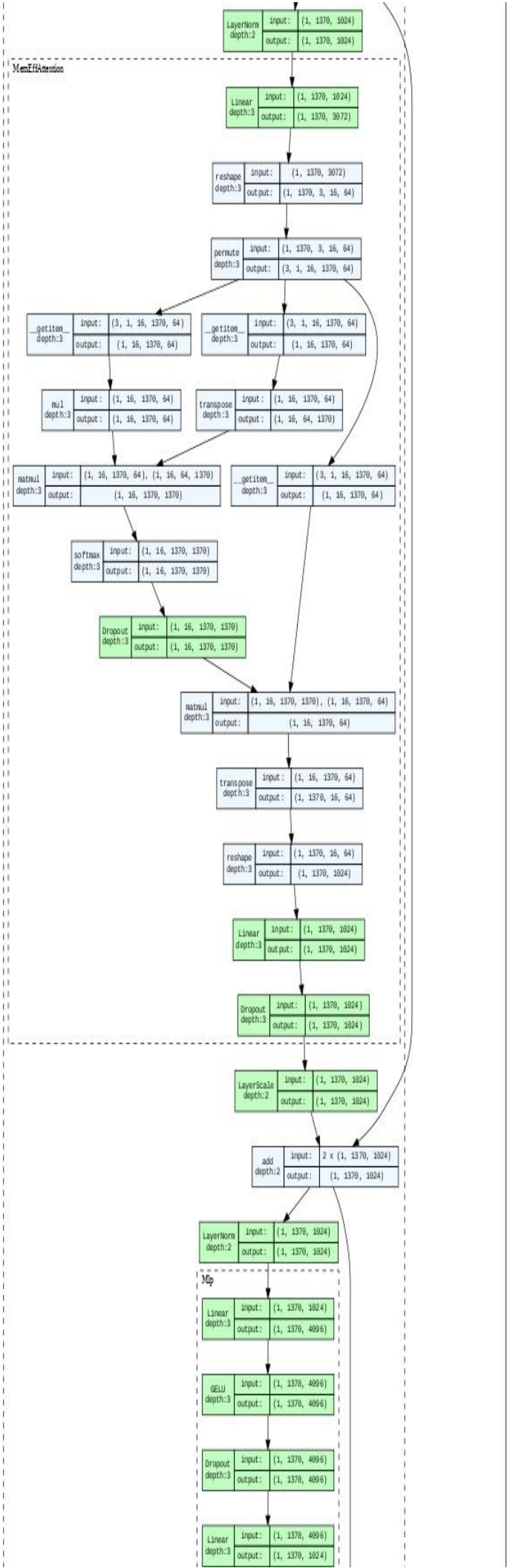

Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)

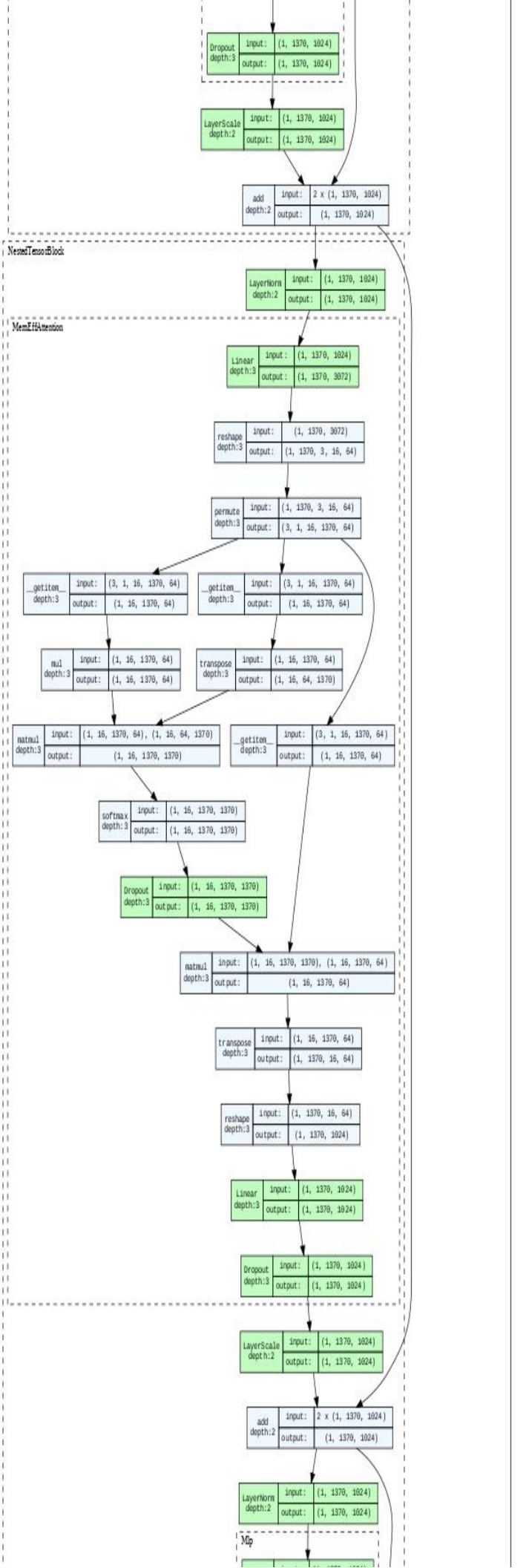

LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose input: (1, 16, 1370, 64)

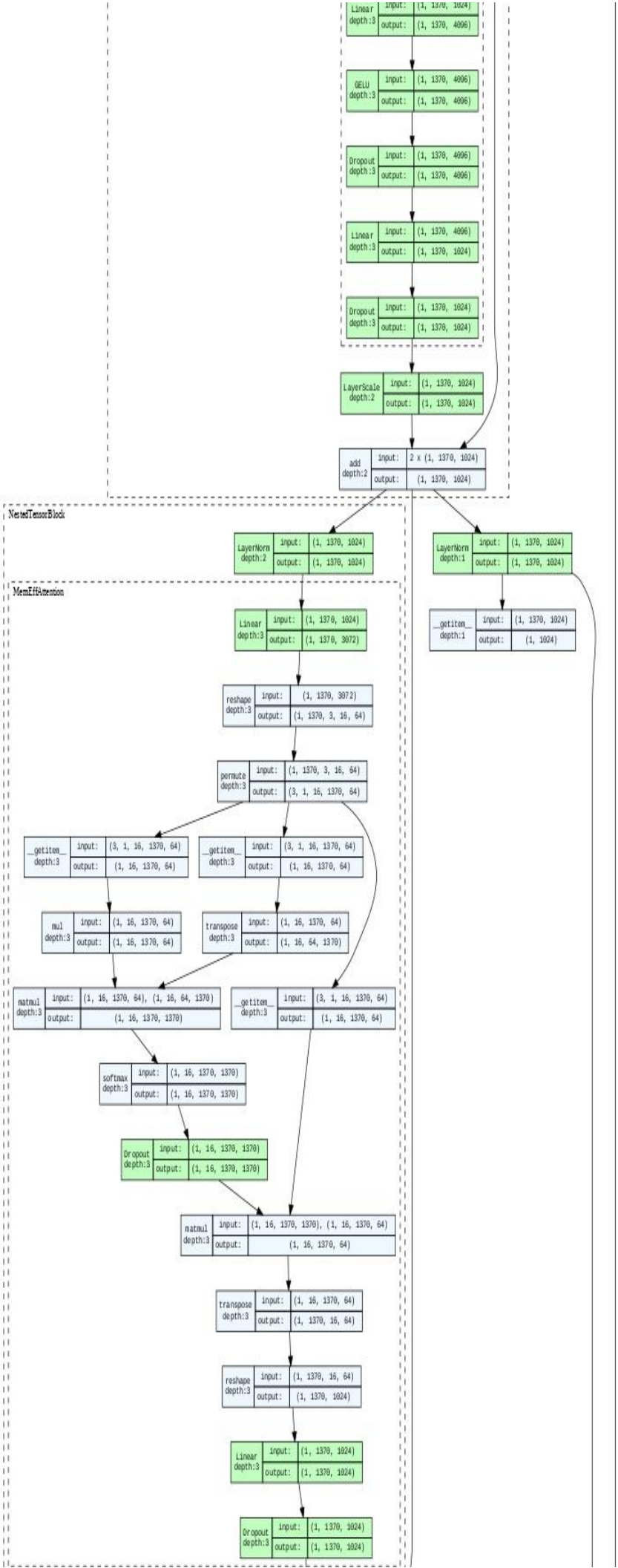

depth:3 output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)

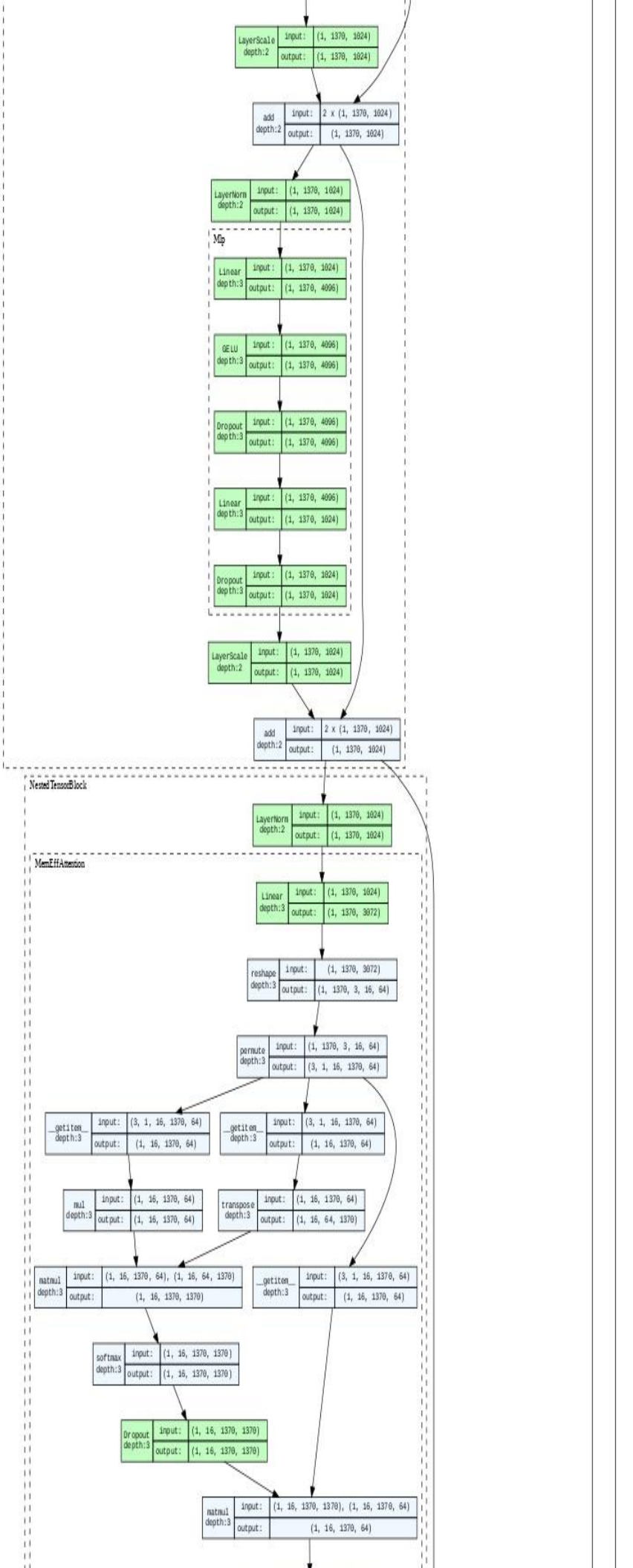

softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute input: (1, 1370, 3, 16, 64)

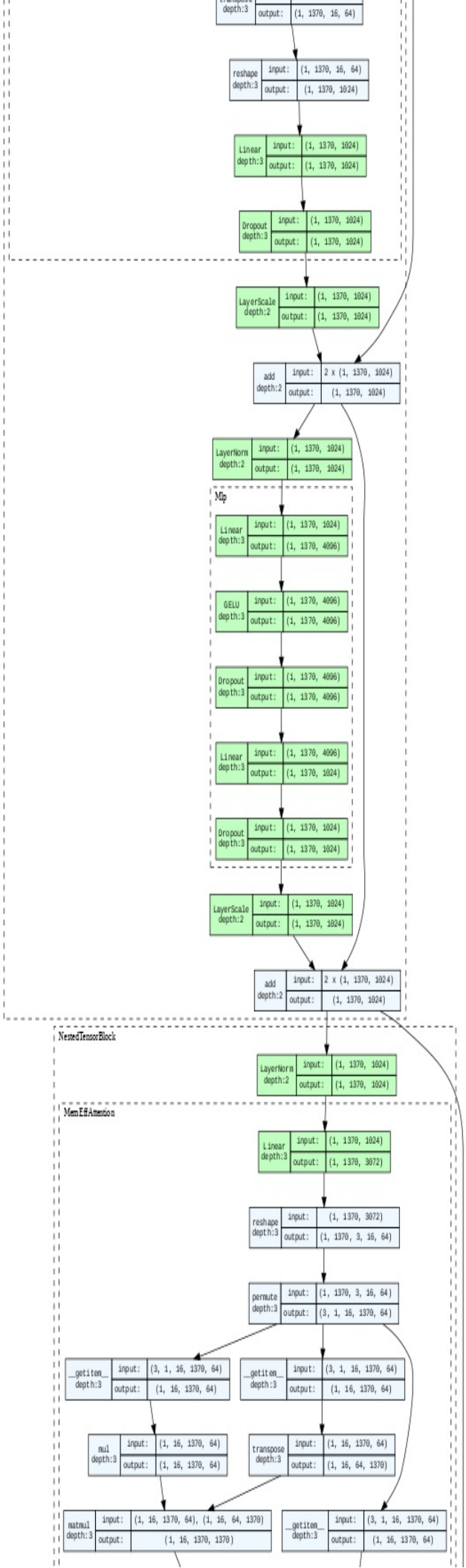

__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock

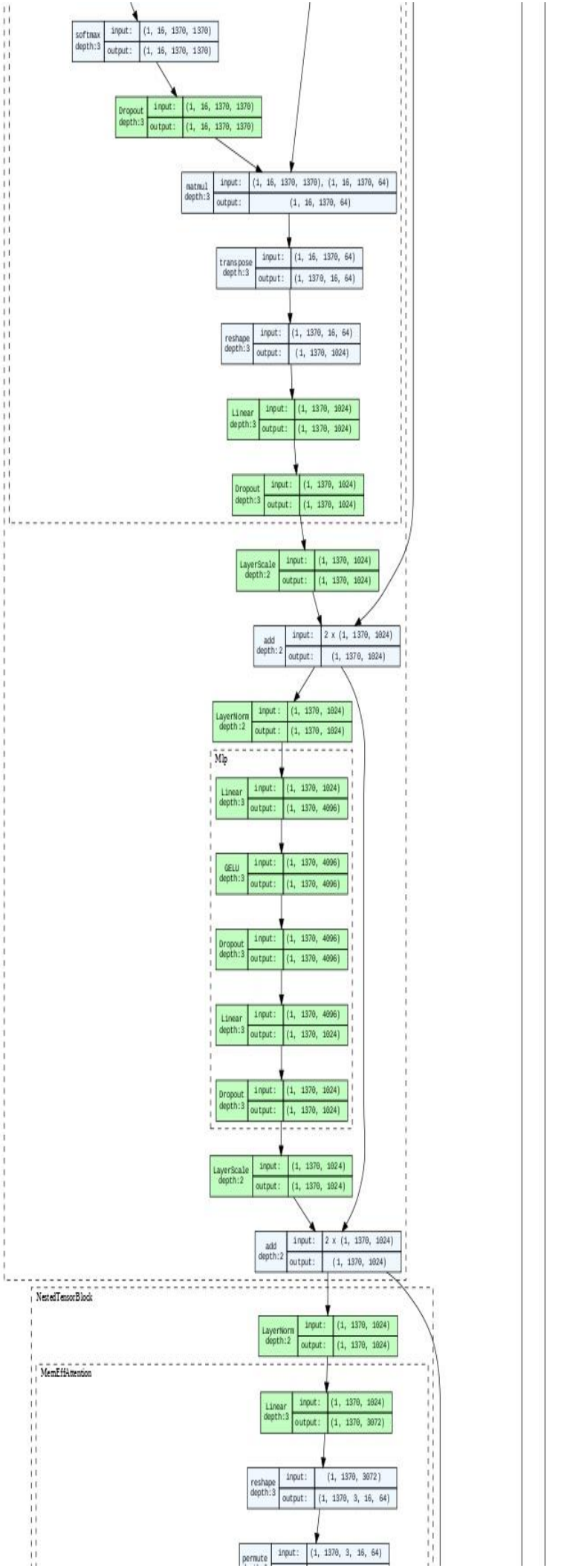

LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)

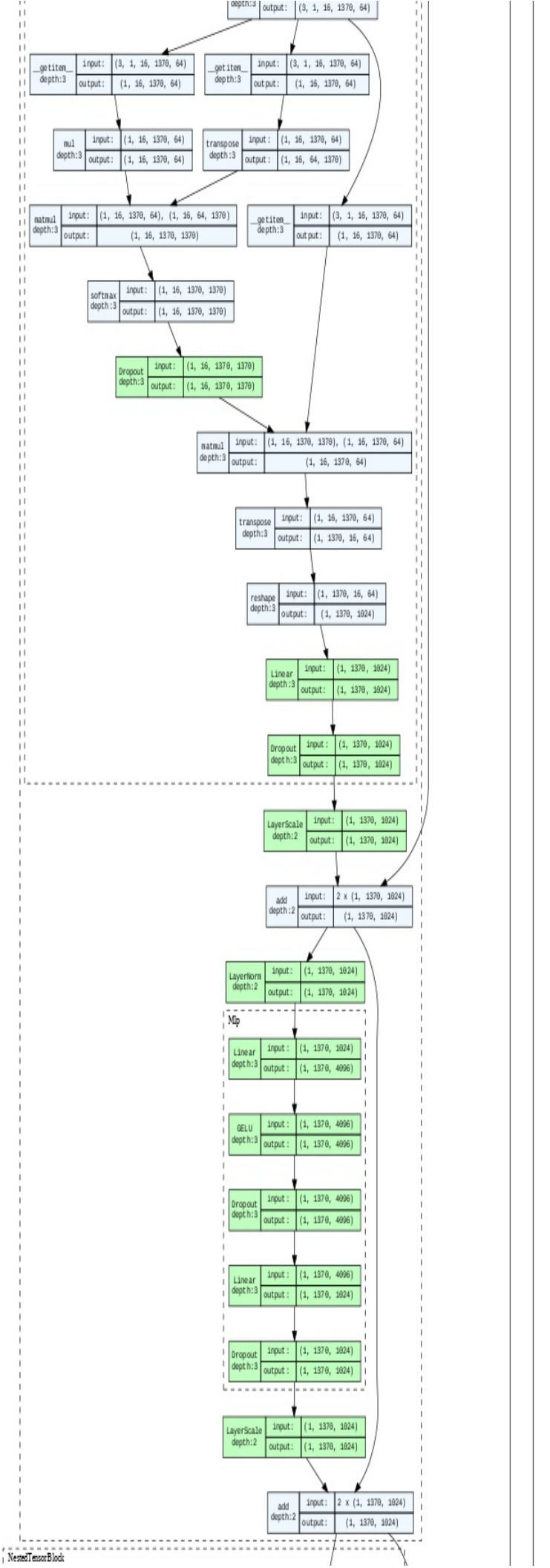

Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp

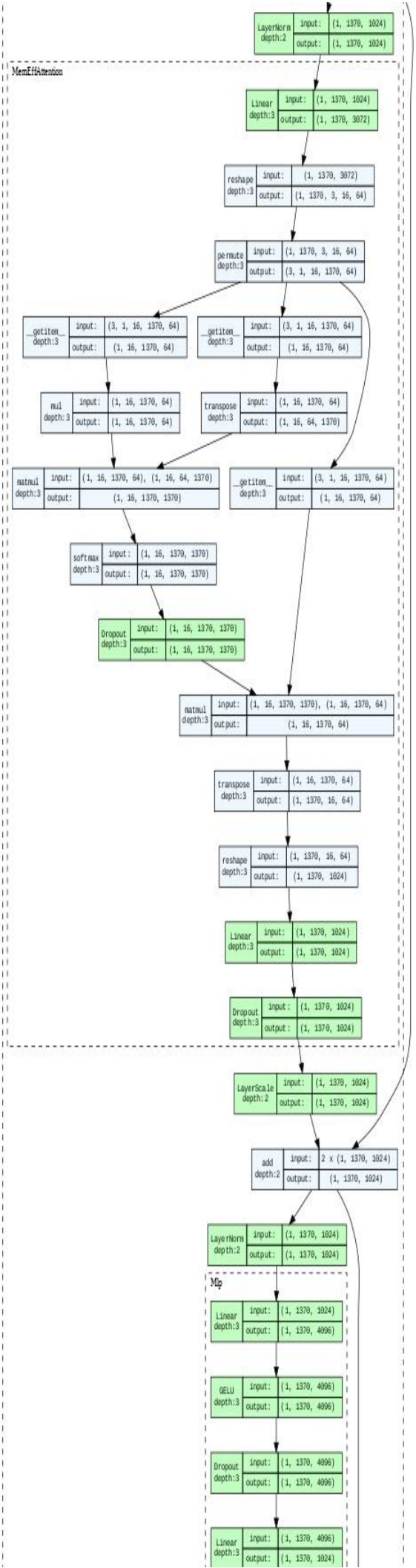

Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:1 input: (1, 1370, 1024) output: (1, 1370, 1024)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)

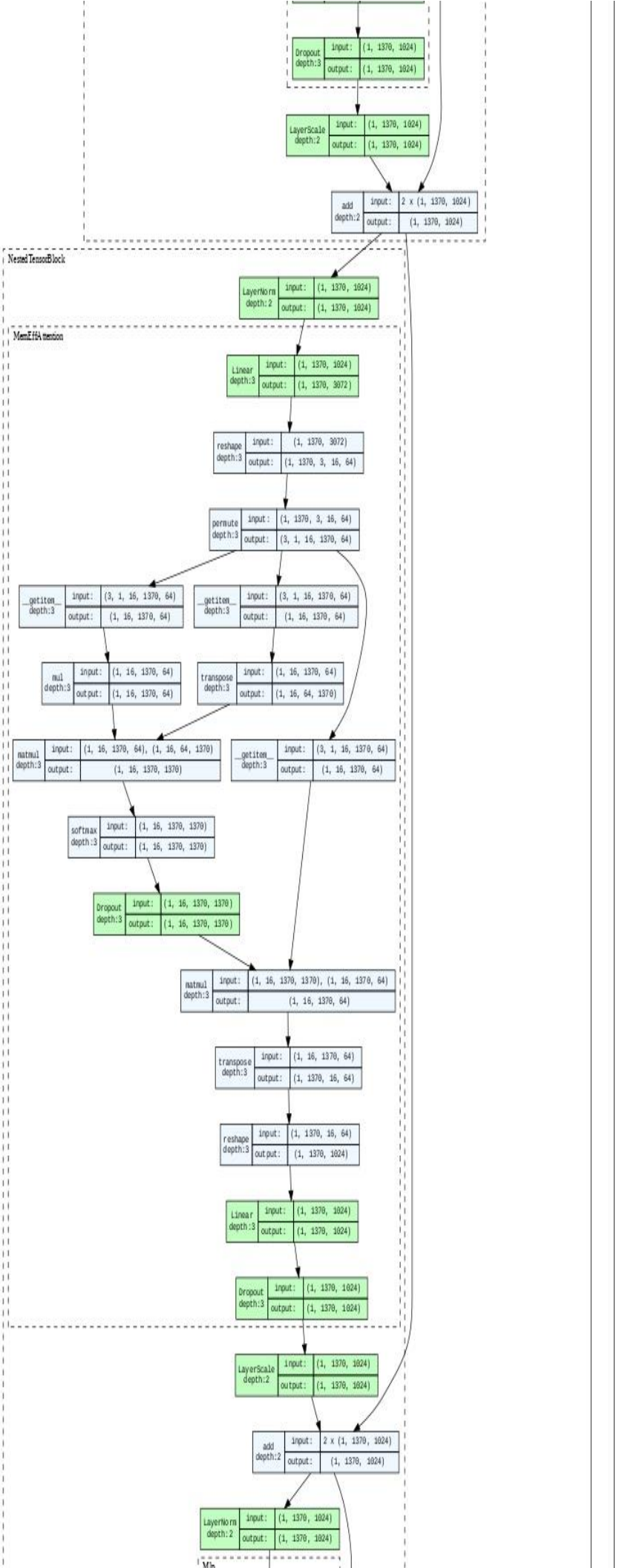

LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)

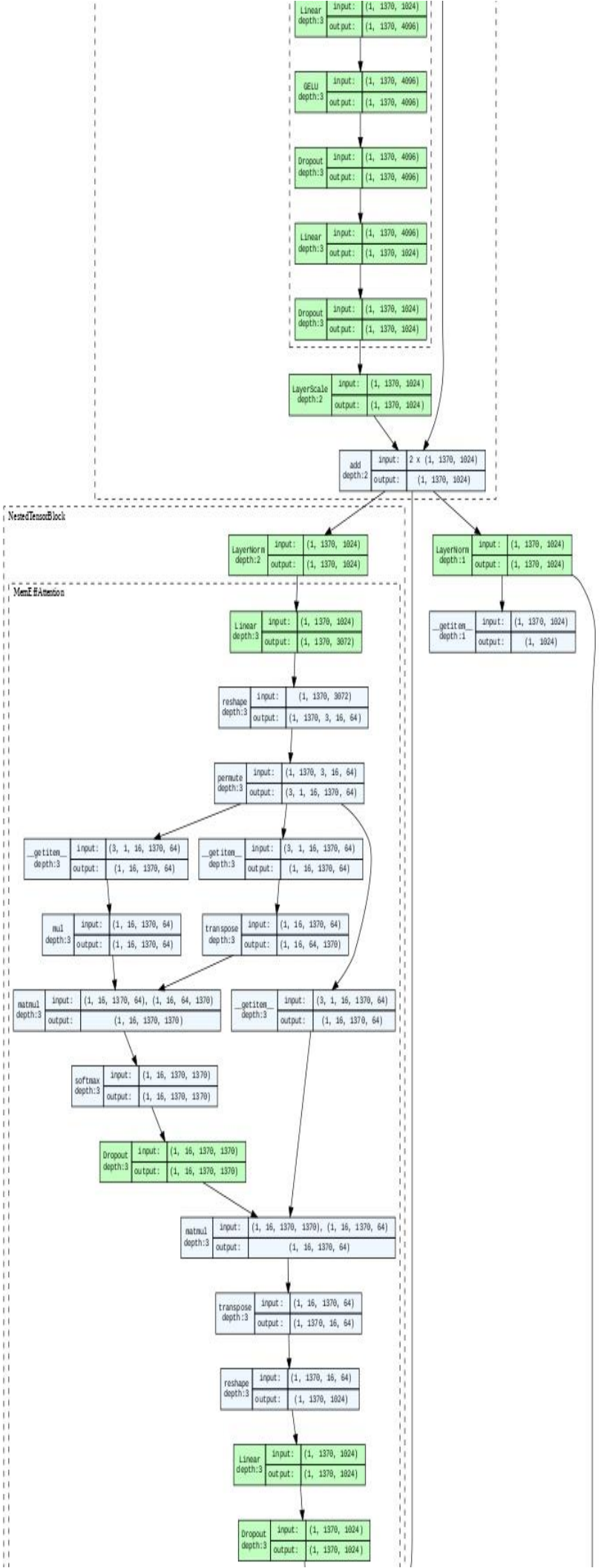

reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)

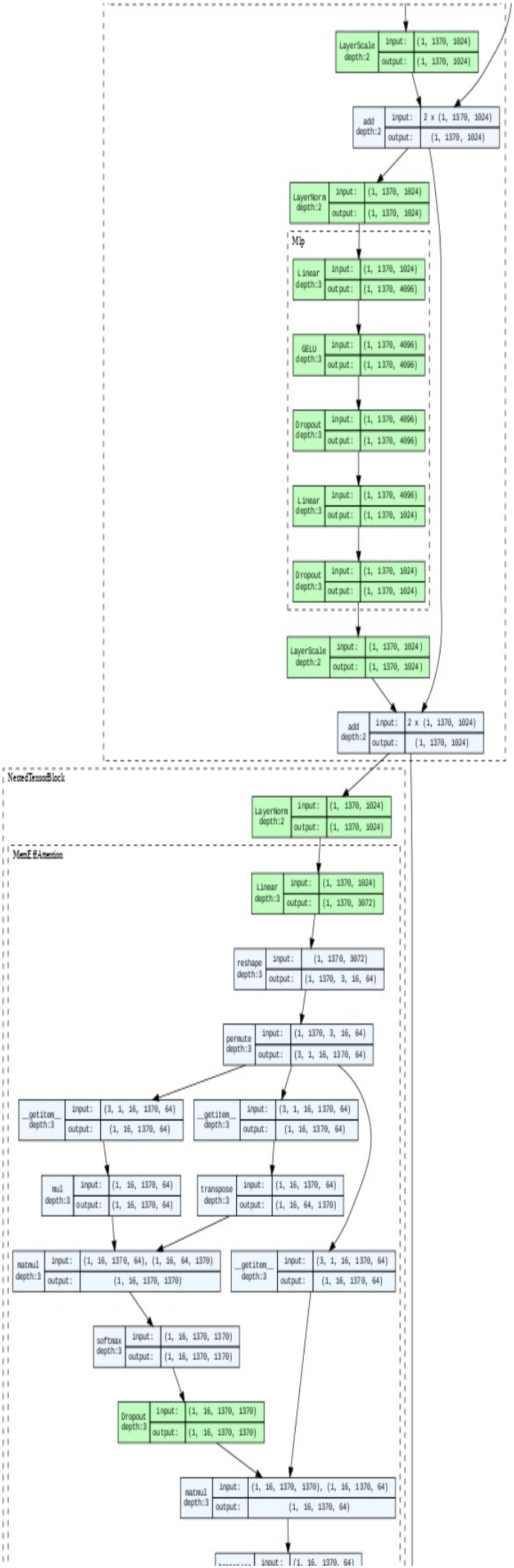

softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute input: (1, 1370, 3, 16, 64)

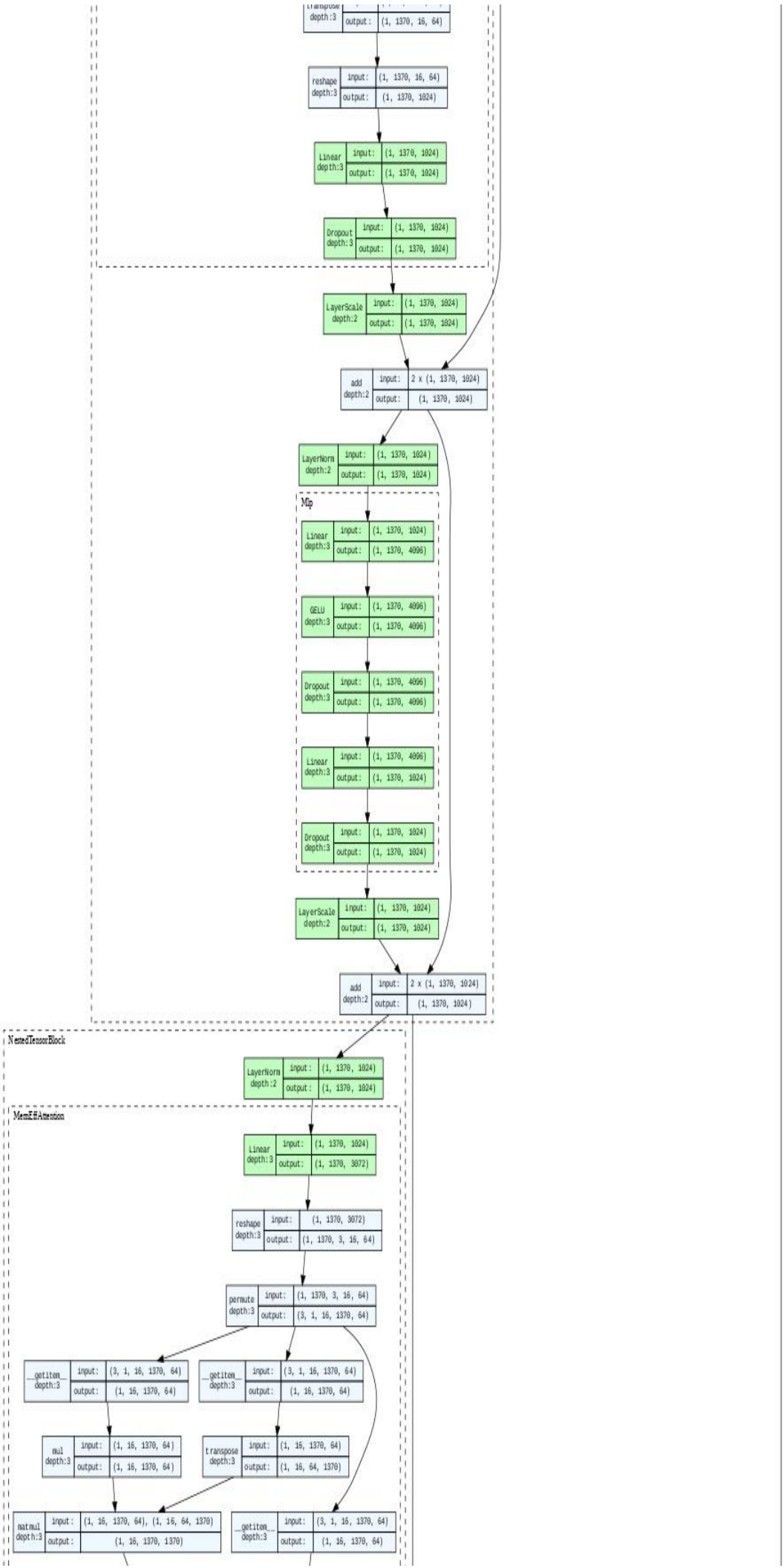

__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock

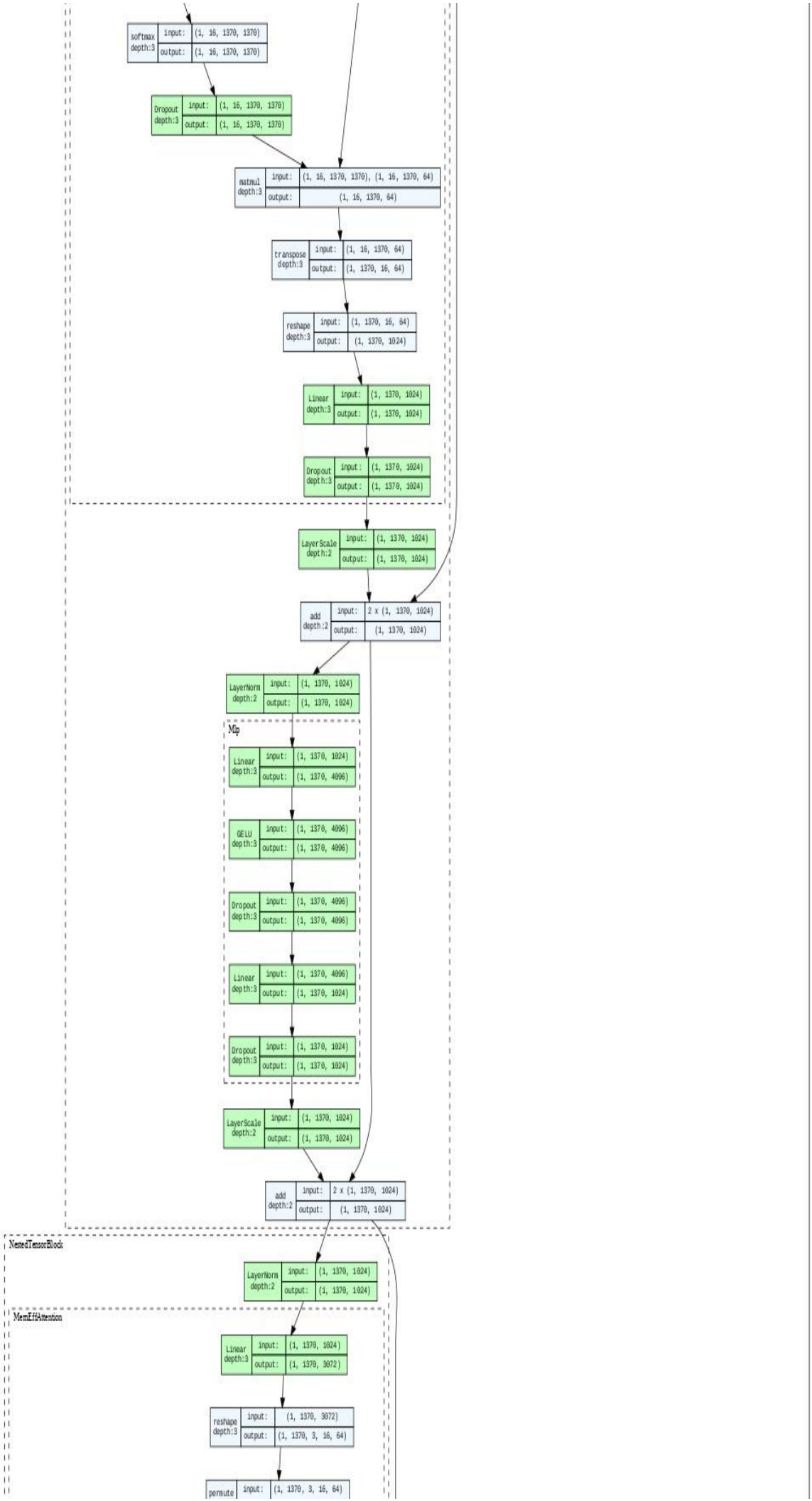

LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)

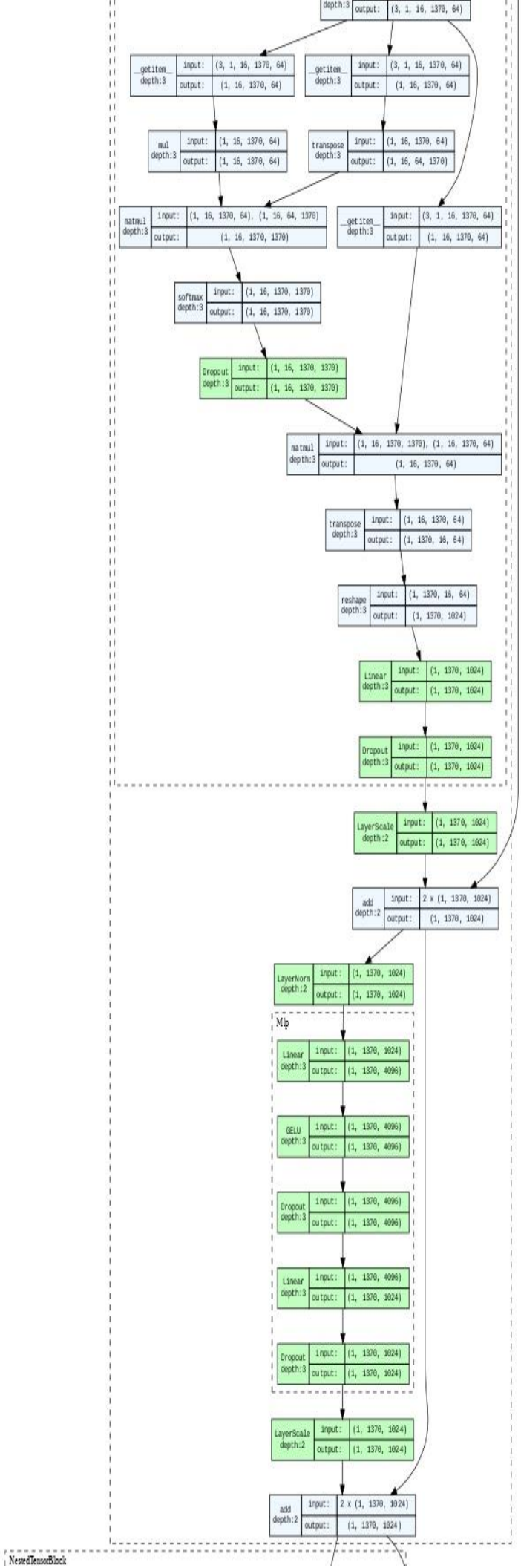

Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)

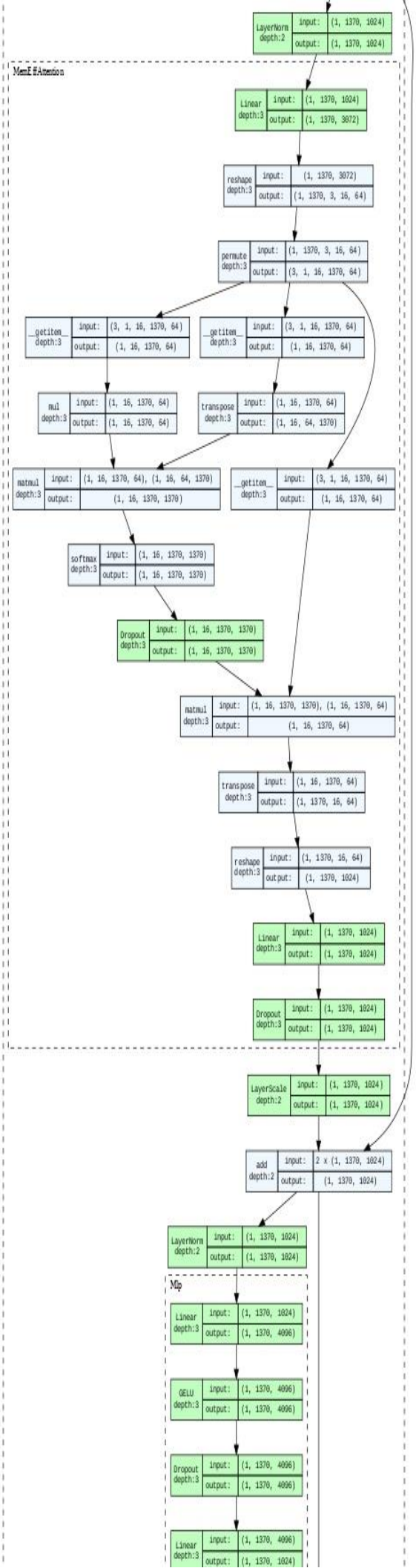

Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:1 input: (1, 1370, 1024) output: (1, 1370, 1024)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)


LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)

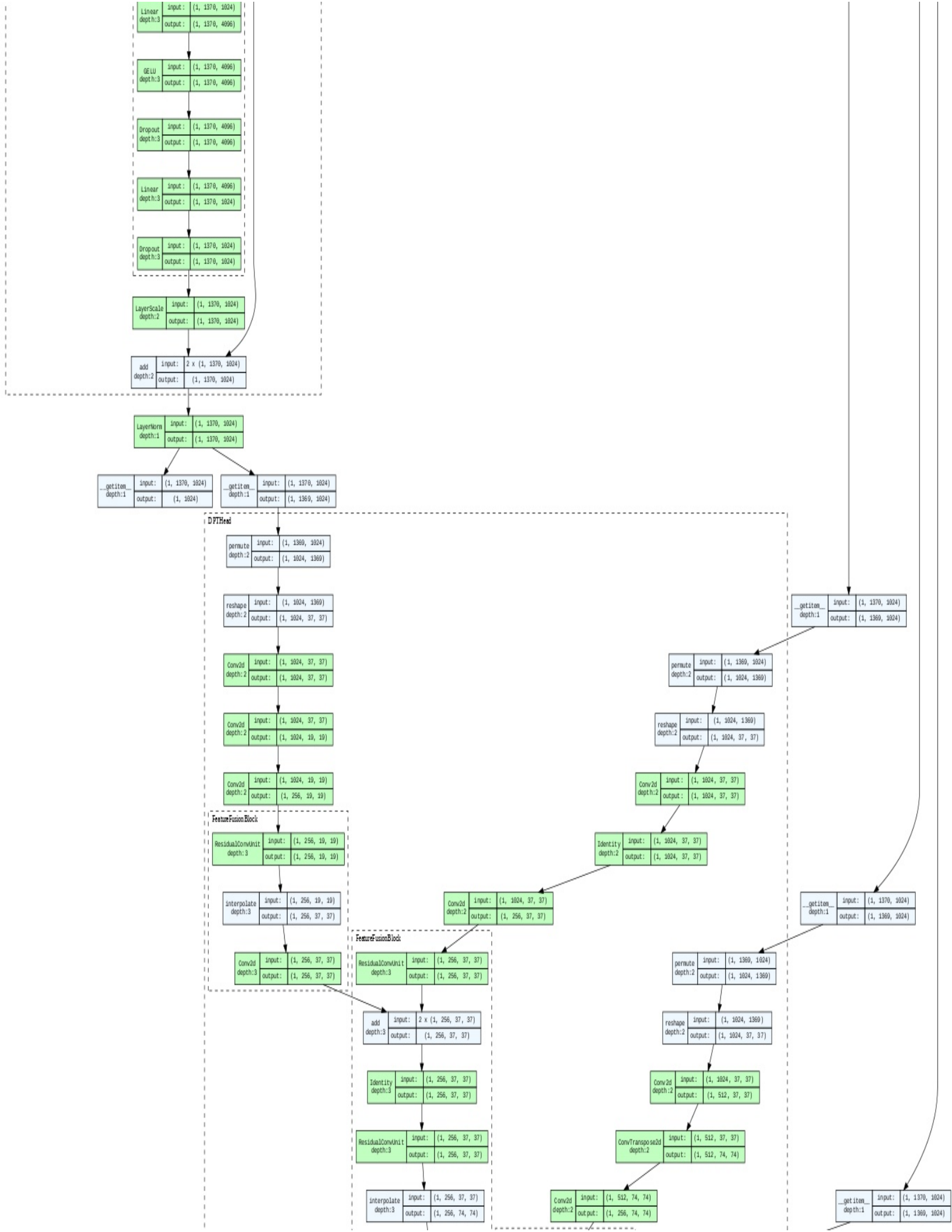

depth:3 output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)

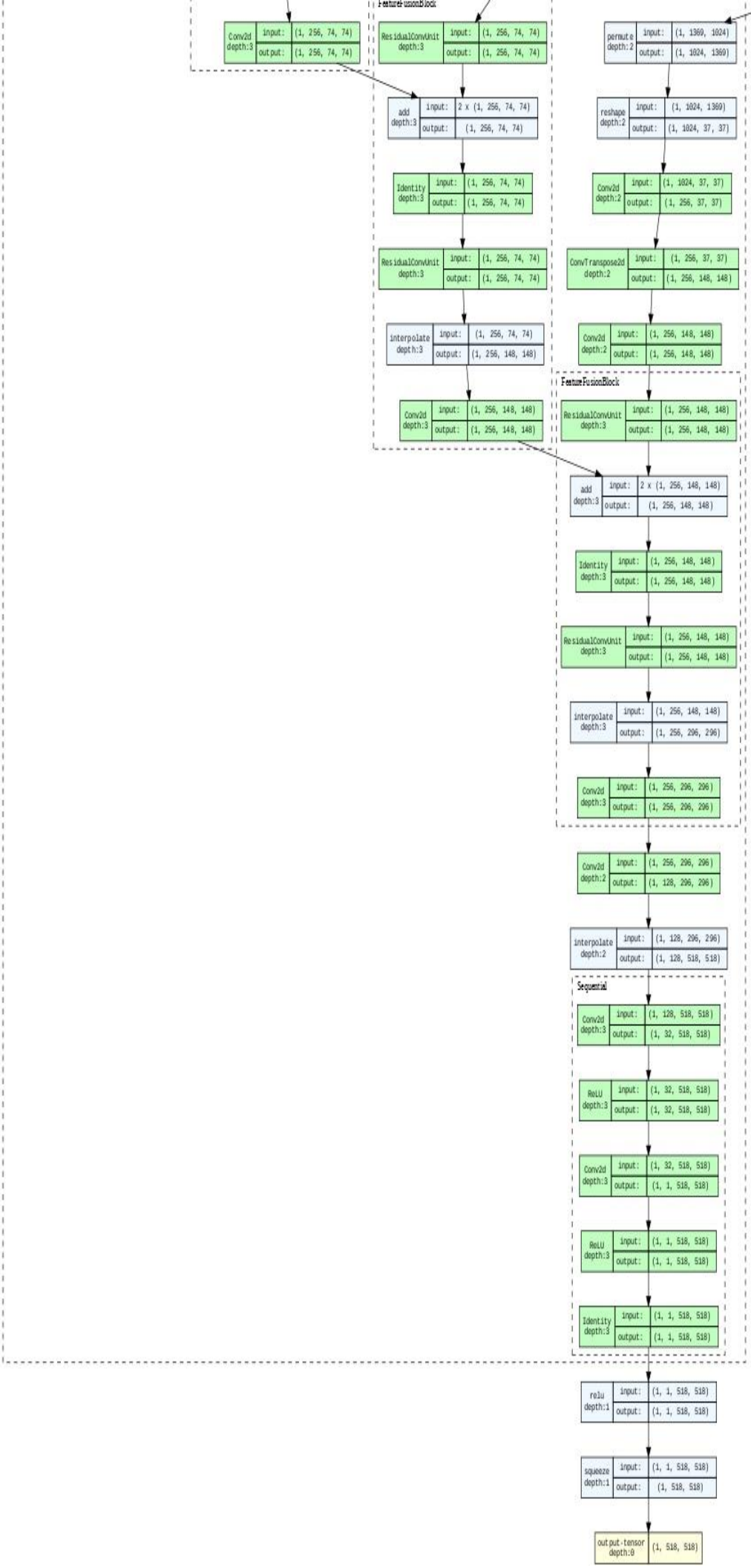

softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute input: (1, 1370, 3, 16, 64)


__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock


LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEfficientAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)


Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
NestedTensorBlock
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
MemEffAttention
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 3072)
reshape depth:3 input: (1, 1370, 3072) output: (1, 1370, 3, 16, 64)
permute depth:3 input: (1, 1370, 3, 16, 64) output: (3, 1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
mul depth:3 input: (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 16, 64, 1370)
matmul depth:3 input: (1, 16, 1370, 64), (1, 16, 64, 1370) output: (1, 16, 1370, 1370)
__getitem__ depth:3 input: (3, 1, 16, 1370, 64) output: (1, 16, 1370, 64)
softmax depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
Dropout depth:3 input: (1, 16, 1370, 1370) output: (1, 16, 1370, 1370)
matmul depth:3 input: (1, 16, 1370, 1370), (1, 16, 1370, 64) output: (1, 16, 1370, 64)
transpose depth:3 input: (1, 16, 1370, 64) output: (1, 1370, 16, 64)
reshape depth:3 input: (1, 1370, 16, 64) output: (1, 1370, 1024)
Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
Mlp


Linear depth:3 input: (1, 1370, 1024) output: (1, 1370, 4096)
GELU depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Dropout depth:3 input: (1, 1370, 4096) output: (1, 1370, 4096)
Linear depth:3 input: (1, 1370, 4096) output: (1, 1370, 1024)
Dropout depth:3 input: (1, 1370, 1024) output: (1, 1370, 1024)
LayerScale depth:2 input: (1, 1370, 1024) output: (1, 1370, 1024)
add depth:2 input: 2 x (1, 1370, 1024) output: (1, 1370, 1024)
LayerNorm depth:1 input: (1, 1370, 1024) output: (1, 1370, 1024)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1024)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1369, 1024)
DPTHead
permute depth:2 input: (1, 1369, 1024) output: (1, 1024, 1369)
reshape depth:2 input: (1, 1024, 1369) output: (1, 1024, 37, 37)
Conv2d depth:2 input: (1, 1024, 37, 37) output: (1, 1024, 37, 37)
Conv2d depth:2 input: (1, 1024, 37, 37) output: (1, 1024, 19, 19)
Conv2d depth:2 input: (1, 1024, 19, 19) output: (1, 256, 19, 19)
FeatureFusionBlock
ResidualConvUnit depth:3 input: (1, 256, 19, 19) output: (1, 256, 19, 19)
interpolate depth:3 input: (1, 256, 19, 19) output: (1, 256, 37, 37)
Conv2d depth:3 input: (1, 256, 37, 37) output: (1, 256, 37, 37)
FeatureFusionBlock
ResidualConvUnit depth:3 input: (1, 256, 37, 37) output: (1, 256, 37, 37)
add depth:3 input: 2 x (1, 256, 37, 37) output: (1, 256, 37, 37)
Identity depth:3 input: (1, 256, 37, 37) output: (1, 256, 37, 37)
ResidualConvUnit depth:3 input: (1, 256, 37, 37) output: (1, 256, 37, 37)
interpolate depth:3 input: (1, 256, 37, 37) output: (1, 256, 74, 74)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1369, 1024)
permute depth:2 input: (1, 1369, 1024) output: (1, 1024, 1369)
reshape depth:2 input: (1, 1024, 1369) output: (1, 1024, 37, 37)
Conv2d depth:2 input: (1, 1024, 37, 37) output: (1, 1024, 37, 37)
Identity depth:2 input: (1, 1024, 37, 37) output: (1, 1024, 37, 37)
Conv2d depth:2 input: (1, 1024, 37, 37) output: (1, 256, 37, 37)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1369, 1024)
permute depth:2 input: (1, 1369, 1024) output: (1, 1024, 1369)
reshape depth:2 input: (1, 1024, 1369) output: (1, 1024, 37, 37)
Conv2d depth:2 input: (1, 1024, 37, 37) output: (1, 512, 37, 37)
ConvTranspose2d depth:2 input: (1, 512, 37, 37) output: (1, 512, 74, 74)
Conv2d depth:2 input: (1, 512, 74, 74) output: (1, 256, 74, 74)
__getitem__ depth:1 input: (1, 1370, 1024) output: (1, 1360, 1024)


Conv2d depth:3 input: (1, 256, 74, 74) output: (1, 256, 74, 74)
FeatureFusionBlock
ResidualConvUnit depth:3 input: (1, 256, 74, 74) output: (1, 256, 74, 74)
permute depth:2 input: (1, 1369, 1024) output: (1, 1024, 1369)
add depth:3 input: 2 x (1, 256, 74, 74) output: (1, 256, 74, 74)
reshape depth:2 input: (1, 1024, 1369) output: (1, 1024, 37, 37)
Identity depth:3 input: (1, 256, 74, 74) output: (1, 256, 74, 74)
Conv2d depth:2 input: (1, 1024, 37, 37) output: (1, 256, 37, 37)
ResidualConvUnit depth:3 input: (1, 256, 74, 74) output: (1, 256, 74, 74)
ConvTranspose2d depth:2 input: (1, 256, 37, 37) output: (1, 256, 148, 148)
interpolate depth:3 input: (1, 256, 74, 74) output: (1, 256, 148, 148)
Conv2d depth:2 input: (1, 256, 148, 148) output: (1, 256, 148, 148)
Conv2d depth:3 input: (1, 256, 148, 148) output: (1, 256, 148, 148)
FeatureFusionBlock
ResidualConvUnit depth:3 input: (1, 256, 148, 148) output: (1, 256, 148, 148)
add depth:3 input: 2 x (1, 256, 148, 148) output: (1, 256, 148, 148)
Identity depth:3 input: (1, 256, 148, 148) output: (1, 256, 148, 148)
ResidualConvUnit depth:3 input: (1, 256, 148, 148) output: (1, 256, 148, 148)
interpolate depth:3 input: (1, 256, 148, 148) output: (1, 256, 296, 296)
Conv2d depth:3 input: (1, 256, 296, 296) output: (1, 256, 296, 296)
Conv2d depth:2 input: (1, 256, 296, 296) output: (1, 128, 296, 296)
interpolate depth:2 input: (1, 128, 296, 296) output: (1, 128, 518, 518)
Sequential
Conv2d depth:3 input: (1, 128, 518, 518) output: (1, 32, 518, 518)
ReLU depth:3 input: (1, 32, 518, 518) output: (1, 32, 518, 518)
Conv2d depth:3 input: (1, 32, 518, 518) output: (1, 1, 518, 518)
ReLU depth:3 input: (1, 1, 518, 518) output: (1, 1, 518, 518)
Identity depth:3 input: (1, 1, 518, 518) output: (1, 1, 518, 518)
relu depth:1 input: (1, 1, 518, 518) output: (1, 1, 518, 518)
squeeze depth:1 input: (1, 1, 518, 518) output: (1, 518, 518)
output-tensor depth:0 (1, 518, 518)