

```
In [1]: from transformers import AutoModelForCausalLM, AutoTokenizer
import torch
from torch.amp import autocast, GradScaler
from pytorch_memlab import MemReporter
import gc

In [2]: model_directory = r"Z:/llmfile/Llama-2-7B-hf/models--meta-llama--Llama-2-7b-hf/snapshots/01c7f73d771dfac7d29232/

model = AutoModelForCausalLM.from_pretrained(model_directory)
tokenizer = AutoTokenizer.from_pretrained(model_directory)

# 设置优化器(Adam)和 GradScaler 用于混合精度
optimizer = torch.optim.AdamW(model.parameters(), lr=1e-5)
scaler = GradScaler()

# 初始输入和标签
input_text = "Hello"
input_ids = tokenizer(input_text, return_tensors="pt").input_ids
labels = input_ids.clone() # 使用输入作为标签

reporter = MemReporter(model)

# 先执行一次前向和反向传播, 确保 scaler 正常初始化
print("==初始化一下==")
model.train()
optimizer.zero_grad()

with autocast("cuda"):
    outputs = model(input_ids, labels=labels)
    loss = outputs.loss

scaler.scale(loss).backward() # 反向传播以初始化 GradScaler

# 执行优化器更新并监控内存
print("==OptimizerMemoryUsage==")
scaler.step(optimizer) # 执行优化器更新
scaler.update() # 更新 scaler 状态
reporter.report() # 记录优化器更新后的内存使用情况
```

Loading checkpoint shards: 0%| | 0/2 [00:00<?, ?it/s]

==初始化一下==

==OptimizerMemoryUsage==

Element type	Size	Used MEM
-----		
Storage on cpu		
Tensor0	(4096, 4096)	64.00M
Tensor1	(1,)	512.00B
Tensor2	(4096, 4096)	64.00M
Tensor3	(4096, 4096)	64.00M
Tensor4	(1,)	512.00B
Tensor5	(11008, 4096)	172.00M
Tensor6	(11008, 4096)	172.00M
Tensor7	(1,)	512.00B
Tensor8	(11008, 4096)	172.00M
Tensor9	(11008, 4096)	172.00M
Tensor10	(1,)	512.00B
Tensor11	(4096, 11008)	172.00M
Tensor12	(4096, 11008)	172.00M
Tensor13	(1,)	512.00B
Tensor14	(4096,)	16.00K
Tensor15	(4096,)	16.00K
Tensor16	(1,)	512.00B
Tensor17	(4096,)	16.00K
Tensor18	(4096,)	16.00K
Tensor19	(1,)	512.00B
Tensor20	(4096, 4096)	64.00M
Tensor21	(4096, 4096)	64.00M
Tensor22	(1,)	512.00B
Tensor23	(4096, 4096)	64.00M
Tensor24	(4096, 4096)	64.00M
Tensor25	(1,)	512.00B
Tensor26	(4096, 4096)	64.00M
Tensor27	(4096, 4096)	64.00M
Tensor28	(1,)	512.00B
Tensor29	(4096, 4096)	64.00M
Tensor30	(4096, 4096)	64.00M
Tensor31	(1,)	512.00B
Tensor32	(11008, 4096)	172.00M
Tensor33	(11008, 4096)	172.00M
Tensor34	(1,)	512.00B
Tensor35	(11008, 4096)	172.00M
Tensor36	(11008, 4096)	172.00M

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

Tensor119	(1,)	512.00B	
Tensor120	(1,)	512.00B	
Tensor121	(1,)	512.00B	
model.layers.30.self_attn.o_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.30.mlp.gate_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.30.mlp.up_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.30.mlp.down_proj.weight.grad	(4096, 11008)	0.00B	
model.layers.30.input_layernorm.weight.grad	(4096,)	0.00B	
model.layers.30.post_attention_layernorm.weight.grad	(4096,)	0.00B	0.00B
model.layers.31.self_attn.q_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.31.self_attn.k_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.31.self_attn.v_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.31.self_attn.o_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.31.mlp.gate_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.31.mlp.up_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.31.mlp.down_proj.weight.grad	(4096, 11008)	0.00B	
model.layers.31.input_layernorm.weight.grad	(4096,)	0.00B	
model.layers.31.post_attention_layernorm.weight.grad	(4096,)	0.00B	0.00B
model.norm.weight.grad	(4096,)	0.00B	
lm_head.weight.grad	(32000, 4096)	0.00B	
Tensor122	(1,)	512.00B	
Tensor123	(32000, 4096)	500.00M	
Tensor124	(32000, 4096)	500.00M	
Tensor125	(1,)	512.00B	
Tensor126	(4096, 4096)	64.00M	
Tensor127	(4096, 4096)	64.00M	
Tensor128	(1,)	512.00B	
Tensor129	(4096, 4096)	64.00M	
Tensor130	(4096, 4096)	64.00M	
Tensor131	(1,)	512.00B	
Tensor132	(4096, 4096)	64.00M	
Tensor133	(4096, 4096)	64.00M	
Tensor134	(1,)	512.00B	
Tensor135	(4096, 4096)	64.00M	
Tensor136	(4096, 4096)	64.00M	
Tensor137	(1,)	512.00B	
Tensor138	(11008, 4096)	172.00M	
Tensor139	(11008, 4096)	172.00M	
Tensor140	(1,)	512.00B	
Tensor141	(11008, 4096)	172.00M	
Tensor142	(11008, 4096)	172.00M	
Tensor143	(1,)	512.00B	
Tensor144	(4096, 11008)	172.00M	
Tensor145	(4096, 11008)	172.00M	
Tensor146	(1,)	512.00B	
Tensor147	(4096,)	16.00K	
Tensor148	(4096,)	16.00K	
Tensor149	(1,)	512.00B	
Tensor150	(4096,)	16.00K	
Tensor151	(4096,)	16.00K	
Tensor152	(1,)	512.00B	
Tensor153	(4096, 4096)	64.00M	
Tensor154	(4096, 4096)	64.00M	
Tensor155	(1,)	512.00B	
Tensor156	(4096, 4096)	64.00M	
Tensor157	(4096, 4096)	64.00M	
Tensor158	(1,)	512.00B	
Tensor159	(4096, 4096)	64.00M	
Tensor160	(4096, 4096)	64.00M	
Tensor161	(1,)	512.00B	
Tensor162	(4096, 4096)	64.00M	
Tensor163	(4096, 4096)	64.00M	
Tensor164	(1,)	512.00B	
Tensor165	(11008, 4096)	172.00M	
Tensor166	(11008, 4096)	172.00M	
Tensor167	(1,)	512.00B	
Tensor168	(11008, 4096)	172.00M	
Tensor169	(11008, 4096)	172.00M	
Tensor170	(1,)	512.00B	
Tensor171	(4096, 11008)	172.00M	
Tensor172	(4096, 11008)	172.00M	
Tensor173	(1,)	512.00B	
Tensor174	(4096,)	16.00K	
Tensor175	(4096,)	16.00K	
Tensor176	(1,)	512.00B	
Tensor177	(4096,)	16.00K	
Tensor178	(4096,)	16.00K	
Tensor179	(1,)	512.00B	
Tensor180	(4096, 4096)	64.00M	
Tensor181	(4096, 4096)	64.00M	
Tensor182	(1,)	512.00B	
Tensor183	(4096, 4096)	64.00M	
Tensor184	(4096, 4096)	64.00M	

Tensor185	(1,)	512.00B
Tensor186	(4096, 4096)	64.00M
Tensor187	(4096, 4096)	64.00M
Tensor188	(1,)	512.00B
Tensor189	(4096, 4096)	64.00M
Tensor190	(4096, 4096)	64.00M
Tensor191	(1,)	512.00B
Tensor192	(11008, 4096)	172.00M
Tensor193	(11008, 4096)	172.00M
Tensor194	(1,)	512.00B
Tensor195	(11008, 4096)	172.00M
Tensor196	(11008, 4096)	172.00M
Tensor197	(1,)	512.00B
Tensor198	(4096, 11008)	172.00M
Tensor199	(4096, 11008)	172.00M
Tensor200	(1,)	512.00B
Tensor201	(4096,)	16.00K
Tensor202	(4096,)	16.00K
Tensor203	(1,)	512.00B
Tensor204	(4096,)	16.00K
Tensor205	(4096,)	16.00K
Tensor206	(1,)	512.00B
Tensor207	(4096, 4096)	64.00M
Tensor208	(4096, 4096)	64.00M
Tensor209	(1,)	512.00B
Tensor210	(4096, 4096)	64.00M
Tensor211	(4096, 4096)	64.00M
Tensor212	(1,)	512.00B
Tensor213	(4096, 4096)	64.00M
Tensor214	(4096, 4096)	64.00M
Tensor215	(1,)	512.00B
Tensor216	(4096, 4096)	64.00M
Tensor217	(4096, 4096)	64.00M
Tensor218	(1,)	512.00B
Tensor219	(11008, 4096)	172.00M
Tensor220	(11008, 4096)	172.00M
Tensor221	(1,)	512.00B
Tensor222	(11008, 4096)	172.00M
Tensor223	(11008, 4096)	172.00M
Tensor224	(1,)	512.00B
Tensor225	(4096, 11008)	172.00M
Tensor226	(4096, 11008)	172.00M
Tensor227	(1,)	512.00B
Tensor228	(4096,)	16.00K
Tensor229	(4096,)	16.00K
Tensor230	(1,)	512.00B
Tensor231	(4096,)	16.00K
Tensor232	(4096,)	16.00K
Tensor233	(1,)	512.00B
Tensor234	(4096, 4096)	64.00M
Tensor235	(4096, 4096)	64.00M
Tensor236	(1,)	512.00B
Tensor237	(4096, 4096)	64.00M
Tensor238	(4096, 4096)	64.00M
Tensor239	(1,)	512.00B
Tensor240	(4096, 4096)	64.00M
Tensor241	(4096, 4096)	64.00M
Tensor242	(1,)	512.00B
Tensor243	(4096, 4096)	64.00M
Tensor244	(4096, 4096)	64.00M
Tensor245	(1,)	512.00B
Tensor246	(11008, 4096)	172.00M
Tensor247	(11008, 4096)	172.00M
Tensor248	(1,)	512.00B
Tensor249	(11008, 4096)	172.00M
Tensor250	(11008, 4096)	172.00M
Tensor251	(1,)	512.00B
Tensor252	(4096, 11008)	172.00M
Tensor253	(4096, 11008)	172.00M
Tensor254	(1,)	512.00B
Tensor255	(4096,)	16.00K
Tensor256	(4096,)	16.00K
Tensor257	(1,)	512.00B
Tensor258	(4096,)	16.00K
Tensor259	(4096,)	16.00K
Tensor260	(1,)	512.00B
Tensor261	(4096, 4096)	64.00M
Tensor262	(4096, 4096)	64.00M
Tensor263	(1,)	512.00B
Tensor264	(4096, 4096)	64.00M
Tensor265	(4096, 4096)	64.00M
Tensor266	(1,)	512.00B
Tensor267	(4096, 4096)	64.00M

Tensor268	(4096, 4096)	64.00M
Tensor269	(1,)	512.00B
Tensor270	(4096, 4096)	64.00M
Tensor271	(4096, 4096)	64.00M
Tensor272	(1,)	512.00B
Tensor273	(11008, 4096)	172.00M
Tensor274	(11008, 4096)	172.00M
Tensor275	(1,)	512.00B
Tensor276	(11008, 4096)	172.00M
Tensor277	(11008, 4096)	172.00M
Tensor278	(1,)	512.00B
Tensor279	(4096, 11008)	172.00M
Tensor280	(4096, 11008)	172.00M
Tensor281	(1,)	512.00B
Tensor282	(4096,)	16.00K
Tensor283	(4096,)	16.00K
Tensor284	(1,)	512.00B
Tensor285	(4096,)	16.00K
Tensor286	(4096,)	16.00K
Tensor287	(1,)	512.00B
Tensor288	(4096, 4096)	64.00M
Tensor289	(4096, 4096)	64.00M
Tensor290	(1,)	512.00B
Tensor291	(4096, 4096)	64.00M
Tensor292	(4096, 4096)	64.00M
Tensor293	(1,)	512.00B
Tensor294	(4096, 4096)	64.00M
Tensor295	(4096, 4096)	64.00M
Tensor296	(1,)	512.00B
Tensor297	(4096, 4096)	64.00M
Tensor298	(4096, 4096)	64.00M
Tensor299	(1,)	512.00B
Tensor300	(11008, 4096)	172.00M
Tensor301	(11008, 4096)	172.00M
Tensor302	(1,)	512.00B
Tensor303	(11008, 4096)	172.00M
Tensor304	(11008, 4096)	172.00M
Tensor305	(1,)	512.00B
Tensor306	(4096, 11008)	172.00M
Tensor307	(4096, 11008)	172.00M
Tensor308	(1,)	512.00B
Tensor309	(4096,)	16.00K
Tensor310	(4096,)	16.00K
Tensor311	(1,)	512.00B
Tensor312	(4096,)	16.00K
Tensor313	(4096,)	16.00K
Tensor314	(1,)	512.00B
Tensor315	(4096, 4096)	64.00M
Tensor316	(4096, 4096)	64.00M
Tensor317	(1,)	512.00B
Tensor318	(4096, 4096)	64.00M
Tensor319	(4096, 4096)	64.00M
Tensor320	(1,)	512.00B
Tensor321	(4096, 4096)	64.00M
Tensor322	(4096, 4096)	64.00M
Tensor323	(1,)	512.00B
Tensor324	(4096, 4096)	64.00M
Tensor325	(4096, 4096)	64.00M
Tensor326	(1,)	512.00B
Tensor327	(11008, 4096)	172.00M
Tensor328	(11008, 4096)	172.00M
Tensor329	(1,)	512.00B
Tensor330	(11008, 4096)	172.00M
Tensor331	(11008, 4096)	172.00M
Tensor332	(1,)	512.00B
Tensor333	(4096, 11008)	172.00M
Tensor334	(4096, 11008)	172.00M
Tensor335	(1,)	512.00B
Tensor336	(4096,)	16.00K
Tensor337	(4096,)	16.00K
Tensor338	(1,)	512.00B
Tensor339	(4096,)	16.00K
Tensor340	(4096,)	16.00K
Tensor341	(1,)	512.00B
Tensor342	(4096, 4096)	64.00M
Tensor343	(4096, 4096)	64.00M
Tensor344	(1,)	512.00B
Tensor345	(4096, 4096)	64.00M
Tensor346	(4096, 4096)	64.00M
Tensor347	(1,)	512.00B
Tensor348	(4096, 4096)	64.00M
Tensor349	(4096, 4096)	64.00M
Tensor350	(1,)	512.00B

Tensor351	(4096, 4096)	64.00M	
Tensor352	(4096, 4096)	64.00M	
Tensor353	(1,)	512.00B	
Tensor354	(11008, 4096)	172.00M	
Tensor355	(11008, 4096)	172.00M	
Tensor356	(1,)	512.00B	
Tensor357	(11008, 4096)	172.00M	
Tensor358	(11008, 4096)	172.00M	
Tensor359	(1,)	512.00B	
Tensor360	(4096, 11008)	172.00M	
Tensor361	(4096, 11008)	172.00M	
Tensor362	(1,)	512.00B	
Tensor363	(4096,)	16.00K	
Tensor364	(4096,)	16.00K	
Tensor365	(1,)	512.00B	
Tensor366	(4096,)	16.00K	
Tensor367	(4096,)	16.00K	
Tensor368	(1,)	512.00B	
Tensor369	(4096, 4096)	64.00M	
Tensor370	(4096, 4096)	64.00M	
Tensor371	(1,)	512.00B	
Tensor372	(4096, 4096)	64.00M	
Tensor373	(4096, 4096)	64.00M	
Tensor374	(1,)	512.00B	
Tensor375	(4096, 4096)	64.00M	
Tensor376	(4096, 4096)	64.00M	
Tensor377	(1,)	512.00B	
Tensor378	(4096, 4096)	64.00M	
Tensor379	(4096, 4096)	64.00M	
Tensor380	(1,)	512.00B	
Tensor381	(11008, 4096)	172.00M	
Tensor382	(11008, 4096)	172.00M	
Tensor383	(1,)	512.00B	
Tensor384	(11008, 4096)	172.00M	
Tensor385	(11008, 4096)	172.00M	
Tensor386	(1,)	512.00B	
Tensor387	(4096, 11008)	172.00M	
Tensor388	(4096, 11008)	172.00M	
Tensor389	(1,)	512.00B	
Tensor390	(4096,)	16.00K	
Tensor391	(4096,)	16.00K	
Tensor392	(1,)	512.00B	
Tensor393	(4096,)	16.00K	
Tensor394	(4096,)	16.00K	
Tensor395	(1,)	512.00B	
Tensor396	(4096, 4096)	64.00M	
Tensor397	(4096, 4096)	64.00M	
Tensor398	(1,)	512.00B	
Tensor399	(4096, 4096)	64.00M	
Tensor400	(4096, 4096)	64.00M	
Tensor401	(1,)	512.00B	
Tensor402	(4096, 4096)	64.00M	
Tensor403	(4096, 4096)	64.00M	
Tensor404	(1,)	512.00B	
Tensor405	(4096, 4096)	64.00M	
Tensor406	(4096, 4096)	64.00M	
Tensor407	(1,)	512.00B	
Tensor408	(11008, 4096)	172.00M	
Tensor409	(11008, 4096)	172.00M	
Tensor410	(1,)	512.00B	
Tensor411	(11008, 4096)	172.00M	
Tensor412	(11008, 4096)	172.00M	
Tensor413	(1,)	512.00B	
Tensor414	(4096, 11008)	172.00M	
Tensor415	(4096, 11008)	172.00M	
Tensor416	(4096,)	16.00K	
Tensor417	(1, 2, 32000)	250.00K	
model.layers.10.input_layernorm.weight.grad	(4096,)	0.00B	
model.layers.10.mlp.down_proj.weight.grad	(4096, 11008)	0.00B	
model.layers.10.mlp.up_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.10.mlp.gate_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.10.self_attn.o_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.10.self_attn.v_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.10.self_attn.k_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.10.self_attn.q_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.9.post_attention_layernorm.weight.grad	(4096,)	0.00B	
model.layers.9.input_layernorm.weight.grad	(4096,)	0.00B	
model.layers.9.mlp.down_proj.weight.grad	(4096, 11008)	0.00B	
model.layers.9.mlp.up_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.9.mlp.gate_proj.weight.grad	(11008, 4096)	0.00B	
model.layers.9.self_attn.o_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.9.self_attn.v_proj.weight.grad	(4096, 4096)	0.00B	
model.layers.9.self_attn.k_proj.weight.grad	(4096, 4096)	0.00B	

[illegible]

Tensor418	(1,)	512.00B
Tensor419	(1, 32, 2, 128)	0.00B
Tensor420	(1, 32, 2, 128)	32.00K
Tensor421	(4096,)	16.00K
Tensor422	(1,)	512.00B
Tensor423	(4096,)	16.00K
Tensor424	(4096,)	16.00K
Tensor425	(1,)	512.00B
Tensor426	(4096, 4096)	64.00M
Tensor427	(4096, 4096)	64.00M
Tensor428	(1,)	512.00B
Tensor429	(4096, 4096)	64.00M
Tensor430	(4096, 4096)	64.00M
Tensor431	(1,)	512.00B
Tensor432	(4096, 4096)	64.00M
Tensor433	(4096, 4096)	64.00M
Tensor434	(1,)	512.00B
Tensor435	(4096, 4096)	64.00M
Tensor436	(4096, 4096)	64.00M
Tensor437	(1,)	512.00B
Tensor438	(11008, 4096)	172.00M
Tensor439	(11008, 4096)	172.00M
Tensor440	(1,)	512.00B
Tensor441	(11008, 4096)	172.00M
Tensor442	(11008, 4096)	172.00M
Tensor443	(1,)	512.00B
Tensor444	(4096, 11008)	172.00M
Tensor445	(4096, 11008)	172.00M
Tensor446	(1,)	512.00B
Tensor447	(4096,)	16.00K
Tensor448	(4096,)	16.00K
Tensor449	(1,)	512.00B
Tensor450	(4096,)	16.00K
Tensor451	(4096,)	16.00K
Tensor452	(1,)	512.00B
Tensor453	(4096, 4096)	64.00M
Tensor454	(4096, 4096)	64.00M
Tensor455	(1,)	512.00B
Tensor456	(4096, 4096)	64.00M
Tensor457	(4096, 4096)	64.00M
Tensor458	(1,)	512.00B
Tensor459	(4096, 4096)	64.00M
Tensor460	(4096, 4096)	64.00M
Tensor461	(1,)	512.00B
Tensor462	(4096, 4096)	64.00M
Tensor463	(4096, 4096)	64.00M
Tensor464	(1,)	512.00B
Tensor465	(11008, 4096)	172.00M
Tensor466	(11008, 4096)	172.00M
Tensor467	(1,)	512.00B
Tensor468	(11008, 4096)	172.00M
Tensor469	(11008, 4096)	172.00M
Tensor470	(1,)	512.00B
Tensor471	(4096, 11008)	172.00M
Tensor472	(4096, 11008)	172.00M
Tensor473	(1,)	512.00B
Tensor474	(4096,)	16.00K
Tensor475	(4096,)	16.00K
Tensor476	(1,)	512.00B
Tensor477	(4096,)	16.00K
Tensor478	(4096,)	16.00K
Tensor479	(1,)	512.00B
Tensor480	(4096, 4096)	64.00M
Tensor481	(4096, 4096)	64.00M
Tensor482	(1,)	512.00B
Tensor483	(4096, 4096)	64.00M
Tensor484	(4096, 4096)	64.00M
Tensor485	(1,)	512.00B
Tensor486	(4096, 4096)	64.00M
Tensor487	(4096, 4096)	64.00M
Tensor488	(1,)	512.00B
Tensor489	(4096, 4096)	64.00M
Tensor490	(4096, 4096)	64.00M
Tensor491	(1,)	512.00B
Tensor492	(11008, 4096)	172.00M
Tensor493	(11008, 4096)	172.00M
Tensor494	(1,)	512.00B
Tensor495	(11008, 4096)	172.00M
Tensor496	(11008, 4096)	172.00M
Tensor497	(1,)	512.00B
Tensor498	(4096, 11008)	172.00M
Tensor499	(4096, 11008)	172.00M
Tensor500	(1,)	512.00B

Tensor501	(4096,)	16.00K
Tensor502	(4096,)	16.00K
Tensor503	(1,)	512.00B
Tensor504	(4096,)	16.00K
Tensor505	(4096,)	16.00K
Tensor506	(1,)	512.00B
Tensor507	(4096, 4096)	64.00M
Tensor508	(4096, 4096)	64.00M
Tensor509	(1,)	512.00B
Tensor510	(4096, 4096)	64.00M
Tensor511	(4096, 4096)	64.00M
Tensor512	(1,)	512.00B
Tensor513	(4096, 4096)	64.00M
Tensor514	(4096, 4096)	64.00M
Tensor515	(1,)	512.00B
Tensor516	(4096, 4096)	64.00M
Tensor517	(4096, 4096)	64.00M
Tensor518	(1,)	512.00B
Tensor519	(11008, 4096)	172.00M
Tensor520	(11008, 4096)	172.00M
Tensor521	(1,)	512.00B
Tensor522	(11008, 4096)	172.00M
Tensor523	(11008, 4096)	172.00M
Tensor524	(1,)	512.00B
Tensor525	(4096, 11008)	172.00M
Tensor526	(4096, 11008)	172.00M
Tensor527	(1,)	512.00B
Tensor528	(4096,)	16.00K
Tensor529	(4096,)	16.00K
Tensor530	(1,)	512.00B
Tensor531	(4096,)	16.00K
Tensor532	(4096,)	16.00K
Tensor533	(1,)	512.00B
Tensor534	(4096, 4096)	64.00M
Tensor535	(4096, 4096)	64.00M
Tensor536	(1,)	512.00B
Tensor537	(4096, 4096)	64.00M
Tensor538	(4096, 4096)	64.00M
Tensor539	(1,)	512.00B
Tensor540	(4096, 4096)	64.00M
Tensor541	(4096, 4096)	64.00M
Tensor542	(1,)	512.00B
Tensor543	(4096, 4096)	64.00M
Tensor544	(4096, 4096)	64.00M
Tensor545	(1,)	512.00B
Tensor546	(11008, 4096)	172.00M
Tensor547	(11008, 4096)	172.00M
Tensor548	(1,)	512.00B
Tensor549	(11008, 4096)	172.00M
Tensor550	(11008, 4096)	172.00M
Tensor551	(1,)	512.00B
Tensor552	(4096, 11008)	172.00M
Tensor553	(4096, 11008)	172.00M
Tensor554	(1,)	512.00B
Tensor555	(4096,)	16.00K
Tensor556	(4096,)	16.00K
Tensor557	(1,)	512.00B
Tensor558	(4096,)	16.00K
Tensor559	(4096,)	16.00K
Tensor560	(1,)	512.00B
Tensor561	(4096, 4096)	64.00M
Tensor562	(4096, 4096)	64.00M
Tensor563	(1,)	512.00B
Tensor564	(4096, 4096)	64.00M
Tensor565	(4096, 4096)	64.00M
Tensor566	(1,)	512.00B
Tensor567	(4096, 4096)	64.00M
Tensor568	(4096, 4096)	64.00M
Tensor569	(1,)	512.00B
Tensor570	(4096, 4096)	64.00M
Tensor571	(4096, 4096)	64.00M
Tensor572	(1,)	512.00B
Tensor573	(11008, 4096)	172.00M
Tensor574	(11008, 4096)	172.00M
Tensor575	(1,)	512.00B
Tensor576	(11008, 4096)	172.00M
Tensor577	(11008, 4096)	172.00M
Tensor578	(1,)	512.00B
Tensor579	(4096, 11008)	172.00M
Tensor580	(4096, 11008)	172.00M
Tensor581	(1,)	512.00B
Tensor582	(4096,)	16.00K
Tensor583	(4096,)	16.00K



Tensor584	(1,)	512.00B
Tensor585	(4096,)	16.00K
Tensor586	(4096,)	16.00K
Tensor587	(1,)	512.00B
Tensor588	(4096, 4096)	64.00M
Tensor589	(4096, 4096)	64.00M
Tensor590	(1,)	512.00B
Tensor591	(4096, 4096)	64.00M
Tensor592	(4096, 4096)	64.00M
Tensor593	(1,)	512.00B
Tensor594	(4096, 4096)	64.00M
Tensor595	(4096, 4096)	64.00M
Tensor596	(1,)	512.00B
Tensor597	(4096, 4096)	64.00M
Tensor598	(4096, 4096)	64.00M
Tensor599	(1,)	512.00B
Tensor600	(11008, 4096)	172.00M
Tensor601	(11008, 4096)	172.00M
Tensor602	(1,)	512.00B
Tensor603	(11008, 4096)	172.00M
Tensor604	(11008, 4096)	172.00M
Tensor605	(1,)	512.00B
Tensor606	(4096, 11008)	172.00M
Tensor607	(4096, 11008)	172.00M
Tensor608	(1,)	512.00B
Tensor609	(4096,)	16.00K
Tensor610	(4096,)	16.00K
Tensor611	(1,)	512.00B
Tensor612	(4096,)	16.00K
Tensor613	(4096,)	16.00K
Tensor614	(1,)	512.00B
Tensor615	(4096, 4096)	64.00M
Tensor616	(4096, 4096)	64.00M
Tensor617	(1,)	512.00B
Tensor618	(4096, 4096)	64.00M
Tensor619	(4096, 4096)	64.00M
Tensor620	(1,)	512.00B
Tensor621	(4096, 4096)	64.00M
Tensor622	(4096, 4096)	64.00M
Tensor623	(1,)	512.00B
Tensor624	(4096, 4096)	64.00M
Tensor625	(4096, 4096)	64.00M
Tensor626	(1,)	512.00B
Tensor627	(11008, 4096)	172.00M
Tensor628	(11008, 4096)	172.00M
Tensor629	(1,)	512.00B
Tensor630	(11008, 4096)	172.00M
Tensor631	(11008, 4096)	172.00M
Tensor632	(1,)	512.00B
Tensor633	(4096, 11008)	172.00M
Tensor634	(4096, 11008)	172.00M
Tensor635	(1,)	512.00B
Tensor636	(4096,)	16.00K
Tensor637	(4096,)	16.00K
Tensor638	(1,)	512.00B
Tensor639	(4096,)	16.00K
Tensor640	(4096,)	16.00K
Tensor641	(1,)	512.00B
Tensor642	(4096, 4096)	64.00M
Tensor643	(4096, 4096)	64.00M
Tensor644	(1,)	512.00B
Tensor645	(4096, 4096)	64.00M
Tensor646	(4096, 4096)	64.00M
Tensor647	(1,)	512.00B
Tensor648	(4096, 4096)	64.00M
Tensor649	(4096, 4096)	64.00M
Tensor650	(1,)	512.00B
Tensor651	(4096, 4096)	64.00M
Tensor652	(4096, 4096)	64.00M
Tensor653	(1,)	512.00B
Tensor654	(11008, 4096)	172.00M
Tensor655	(11008, 4096)	172.00M
Tensor656	(1,)	512.00B
Tensor657	(11008, 4096)	172.00M
Tensor658	(11008, 4096)	172.00M
Tensor659	(1,)	512.00B
Tensor660	(4096, 11008)	172.00M
Tensor661	(4096, 11008)	172.00M
Tensor662	(1,)	512.00B
Tensor663	(4096,)	16.00K
Tensor664	(4096,)	16.00K
Tensor665	(1,)	512.00B
Tensor666	(4096,)	16.00K

Tensor667	(4096,)	16.00K
Tensor668	(1,)	512.00B
Tensor669	(4096, 4096)	64.00M
Tensor670	(4096, 4096)	64.00M
Tensor671	(1,)	512.00B
Tensor672	(4096, 4096)	64.00M
Tensor673	(4096, 4096)	64.00M
Tensor674	(1,)	512.00B
Tensor675	(4096, 4096)	64.00M
Tensor676	(4096, 4096)	64.00M
Tensor677	(1,)	512.00B
Tensor678	(4096, 4096)	64.00M
Tensor679	(4096, 4096)	64.00M
Tensor680	(1,)	512.00B
Tensor681	(11008, 4096)	172.00M
Tensor682	(11008, 4096)	172.00M
Tensor683	(1,)	512.00B
Tensor684	(11008, 4096)	172.00M
Tensor685	(11008, 4096)	172.00M
Tensor686	(1,)	512.00B
Tensor687	(4096, 11008)	172.00M
Tensor688	(4096, 11008)	172.00M
Tensor689	(1,)	512.00B
Tensor690	(4096,)	16.00K
Tensor691	(4096,)	16.00K
Tensor692	(1,)	512.00B
Tensor693	(4096,)	16.00K
Tensor694	(4096,)	16.00K
Tensor695	(1,)	512.00B
Tensor696	(4096, 4096)	64.00M
Tensor697	(4096, 4096)	64.00M
Tensor698	(1,)	512.00B
Tensor699	(4096, 4096)	64.00M
Tensor700	(4096, 4096)	64.00M
Tensor701	(1,)	512.00B
Tensor702	(4096, 4096)	64.00M
Tensor703	(4096, 4096)	64.00M
Tensor704	(1,)	512.00B
Tensor705	(4096, 4096)	64.00M
Tensor706	(4096, 4096)	64.00M
Tensor707	(1,)	512.00B
Tensor708	(11008, 4096)	172.00M
Tensor709	(11008, 4096)	172.00M
Tensor710	(1,)	512.00B
Tensor711	(11008, 4096)	172.00M
Tensor712	(11008, 4096)	172.00M
Tensor713	(1,)	512.00B
Tensor714	(4096, 11008)	172.00M
Tensor715	(4096, 11008)	172.00M
Tensor716	(1,)	512.00B
Tensor717	(4096,)	16.00K
Tensor718	(4096,)	16.00K
Tensor719	(1,)	512.00B
Tensor720	(4096,)	16.00K
Tensor721	(4096,)	16.00K
Tensor722	(1,)	512.00B
Tensor723	(4096, 4096)	64.00M
Tensor724	(4096, 4096)	64.00M
Tensor725	(1,)	512.00B
Tensor726	(4096, 4096)	64.00M
Tensor727	(4096, 4096)	64.00M
Tensor728	(1,)	512.00B
Tensor729	(4096, 4096)	64.00M
Tensor730	(4096, 4096)	64.00M
Tensor731	(1,)	512.00B
Tensor732	(4096, 4096)	64.00M
Tensor733	(4096, 4096)	64.00M
Tensor734	(1,)	512.00B
Tensor735	(11008, 4096)	172.00M
Tensor736	(11008, 4096)	172.00M
Tensor737	(1,)	512.00B
Tensor738	(11008, 4096)	172.00M
Tensor739	(11008, 4096)	172.00M
Tensor740	(1,)	512.00B
Tensor741	(4096, 11008)	172.00M
Tensor742	(4096, 11008)	172.00M
Tensor743	(1,)	512.00B
Tensor744	(4096,)	16.00K
Tensor745	(4096,)	16.00K
Tensor746	(1,)	512.00B
Tensor747	(4096,)	16.00K
Tensor748	(4096,)	16.00K
Tensor749	(1,)	512.00B

Tensor750	(4096, 4096)	64.00M
Tensor751	(4096, 4096)	64.00M
Tensor752	(1,)	512.00B
Tensor753	(4096, 4096)	64.00M
Tensor754	(4096, 4096)	64.00M
Tensor755	(1,)	512.00B
Tensor756	(4096, 4096)	64.00M
Tensor757	(4096, 4096)	64.00M
Tensor758	(1,)	512.00B
Tensor759	(4096, 4096)	64.00M
Tensor760	(4096, 4096)	64.00M
Tensor761	(1,)	512.00B
Tensor762	(11008, 4096)	172.00M
Tensor763	(11008, 4096)	172.00M
Tensor764	(1,)	512.00B
Tensor765	(11008, 4096)	172.00M
Tensor766	(11008, 4096)	172.00M
Tensor767	(1,)	512.00B
Tensor768	(4096, 11008)	172.00M
Tensor769	(4096, 11008)	172.00M
Tensor770	(1,)	512.00B
Tensor771	(4096,)	16.00K
Tensor772	(4096,)	16.00K
Tensor773	(1,)	512.00B
Tensor774	(4096,)	16.00K
Tensor775	(4096,)	16.00K
Tensor776	(1,)	512.00B
Tensor777	(4096, 4096)	64.00M
Tensor778	(4096, 4096)	64.00M
Tensor779	(1,)	512.00B
Tensor780	(4096, 4096)	64.00M
Tensor781	(4096, 4096)	64.00M
Tensor782	(1,)	512.00B
Tensor783	(4096, 4096)	64.00M
Tensor784	(4096, 4096)	64.00M
Tensor785	(1,)	512.00B
Tensor786	(4096, 4096)	64.00M
Tensor787	(4096, 4096)	64.00M
Tensor788	(1,)	512.00B
Tensor789	(11008, 4096)	172.00M
Tensor790	(11008, 4096)	172.00M
Tensor791	(1,)	512.00B
Tensor792	(11008, 4096)	172.00M
Tensor793	(11008, 4096)	172.00M
Tensor794	(1,)	512.00B
Tensor795	(4096, 11008)	172.00M
Tensor796	(4096, 11008)	172.00M
Tensor797	(1,)	512.00B
Tensor798	(4096,)	16.00K
Tensor799	(4096,)	16.00K
Tensor800	(1,)	512.00B
Tensor801	(4096,)	16.00K
Tensor802	(4096,)	16.00K
Tensor803	(1,)	512.00B
Tensor804	(4096, 4096)	64.00M
Tensor805	(4096, 4096)	64.00M
Tensor806	(1,)	512.00B
Tensor807	(4096, 4096)	64.00M
Tensor808	(4096, 4096)	64.00M
Tensor809	(1,)	512.00B
Tensor810	(4096, 4096)	64.00M
Tensor811	(4096, 4096)	64.00M
Tensor812	(1,)	512.00B
Tensor813	(4096, 4096)	64.00M
Tensor814	(4096, 4096)	64.00M
Tensor815	(1,)	512.00B
Tensor816	(11008, 4096)	172.00M
Tensor817	(11008, 4096)	172.00M
Tensor818	(1,)	512.00B
Tensor819	(11008, 4096)	172.00M
Tensor820	(11008, 4096)	172.00M
Tensor821	(1,)	512.00B
Tensor822	(4096, 11008)	172.00M
Tensor823	(4096, 11008)	172.00M
Tensor824	(1,)	512.00B
Tensor825	(4096,)	16.00K
Tensor826	(4096,)	16.00K
Tensor827	(1,)	512.00B
Tensor828	(4096,)	16.00K
Tensor829	(4096,)	16.00K
Tensor830	(1,)	512.00B
Tensor831	(4096, 4096)	64.00M
Tensor832	(4096, 4096)	64.00M

Tensor833	(1,)	512.00B
Tensor834	(4096, 4096)	64.00M
Tensor835	(4096, 4096)	64.00M
Tensor836	(1,)	512.00B
Tensor837	(4096, 4096)	64.00M
Tensor838	(4096, 4096)	64.00M
Tensor839	(1,)	512.00B
Tensor840	(4096, 4096)	64.00M
Tensor841	(4096, 4096)	64.00M
Tensor842	(1,)	512.00B
Tensor843	(11008, 4096)	172.00M
Tensor844	(11008, 4096)	172.00M
Tensor845	(1,)	512.00B
Tensor846	(11008, 4096)	172.00M
Tensor847	(11008, 4096)	172.00M
Tensor848	(1,)	512.00B
Tensor849	(4096, 11008)	172.00M
Tensor850	(4096, 11008)	172.00M
Tensor851	(1,)	512.00B
Tensor852	(4096,)	16.00K
Tensor853	(4096,)	16.00K
Tensor854	(1,)	512.00B
Tensor855	(4096,)	16.00K
Tensor856	(4096,)	16.00K
Tensor857	(1,)	512.00B
Tensor858	(4096, 4096)	64.00M
Tensor859	(4096, 4096)	64.00M
Tensor860	(1,)	512.00B
Tensor861	(4096, 4096)	64.00M
Tensor862	(4096, 4096)	64.00M
Tensor863	(1,)	512.00B
Tensor864	(4096, 4096)	64.00M
Tensor865	(4096, 4096)	64.00M
Tensor866	(1,)	512.00B
Tensor867	(4096, 4096)	64.00M
Tensor868	(4096, 4096)	64.00M
Tensor869	(1,)	512.00B
Tensor870	(11008, 4096)	172.00M
Tensor871	(11008, 4096)	172.00M
Tensor872	(1,)	512.00B
Tensor873	(11008, 4096)	172.00M
Tensor874	(11008, 4096)	172.00M
Tensor875	(1,)	512.00B
Tensor876	(4096, 11008)	172.00M
Tensor877	(4096, 11008)	172.00M
Tensor878	(1,)	512.00B
Tensor879	(4096,)	16.00K
Tensor880	(4096,)	16.00K
Tensor881	(1,)	512.00B
Tensor882	(4096,)	16.00K
Tensor883	(4096,)	16.00K
Tensor884	(1,)	512.00B
Tensor885	(4096, 4096)	64.00M
Tensor886	(4096, 4096)	64.00M
Tensor887	(1,)	512.00B
Tensor888	(4096, 4096)	64.00M
Tensor889	(4096, 4096)	64.00M
Tensor890	(1,)	512.00B
Tensor891	(4096, 4096)	64.00M
Tensor892	(4096, 4096)	64.00M
Tensor893	(1,)	512.00B
Tensor894	(4096, 4096)	64.00M
Tensor895	(4096, 4096)	64.00M
Tensor896	(1,)	512.00B
Tensor897	(11008, 4096)	172.00M
Tensor898	(11008, 4096)	172.00M
Tensor899	(1,)	512.00B
Tensor900	(11008, 4096)	172.00M
Tensor901	(11008, 4096)	172.00M
Tensor902	(1,)	512.00B
Tensor903	(4096, 11008)	172.00M
Tensor904	(4096, 11008)	172.00M
Tensor905	(1,)	512.00B
Tensor906	(4096,)	16.00K
Tensor907	(4096,)	16.00K
Tensor908	(1,)	512.00B
Tensor909	(4096,)	16.00K
Tensor910	(4096,)	16.00K
Tensor911	(1,)	512.00B
Tensor912	(4096, 4096)	64.00M
Tensor913	(4096, 4096)	64.00M
Tensor914	(1,)	512.00B
Tensor915	(4096, 4096)	64.00M

[illegible]

[illegible]

[illegible]

Tensor990	(1, 32, 2, 128)	32.00K
Tensor991	(1, 32, 2, 128)	0.00B
Tensor992	(1, 32, 2, 128)	32.00K
Tensor993	(1, 32, 2, 128)	0.00B
Tensor994	(1, 32, 2, 128)	32.00K
Tensor995	(1, 32, 2, 128)	0.00B
Tensor996	(1, 32, 2, 128)	32.00K
Tensor997	(1, 32, 2, 128)	0.00B
Tensor998	(1, 32, 2, 128)	32.00K
Tensor999	(1, 32, 2, 128)	0.00B
Tensor1000	(1, 32, 2, 128)	32.00K
Tensor1001	(1, 32, 2, 128)	0.00B
Tensor1002	(1, 32, 2, 128)	32.00K
Tensor1003	(1, 32, 2, 128)	0.00B
Tensor1004	(1, 32, 2, 128)	32.00K
Tensor1005	(1, 32, 2, 128)	0.00B
Tensor1006	(1, 32, 2, 128)	32.00K
Tensor1007	(1,)	512.00B

-----  
Total Tensors: 33692930922      Used Memory: 100.41G  
-----

C:\Users\xxc13\anaconda3\Lib\site-packages\pytorch\_memlab\mem\_reporter.py:65: FutureWarning: `torch.distributed.reduce\_op` is deprecated, please use `torch.distributed.ReduceOp` instead  
  tensors = [obj for obj in objects if isinstance(obj, torch.Tensor)]  
C:\Users\xxc13\anaconda3\Lib\site-packages\pytorch\_memlab\mem\_reporter.py:95: UserWarning: TypedStorage is deprecated. It will be removed in the future and UntypedStorage will be the only storage class. This should only matter to you if you are using storages directly. To access UntypedStorage directly, use tensor.untyped\_storage() instead of tensor.storage()  
  fact\_numel = tensor.storage().size()

```
In [3]: # 清理内存
del outputs, loss
gc.collect()
```

Out[3]: 46

In [ ]:

In [ ]: