Training Loop with Compressive Memory Management

Critical:
Memory state persists
across batches but
gradients do not flow
through history

Start Epoch
compressed_memories = None

Load Batch
32 × 512 bytes
from DataLoader

Forward Pass
logits, memories = model(
batch,
compressed_memories
)

Updated Memories
8 layers × memory states
[batch, mem_len, 512]

Compute Loss
Cross-Entropy
$BPD = -\Sigma \log_2 p(x\_i|x\_{<i}) / N$

Backward Pass
loss.backward()
Compute gradients

Gradient Clipping
max_norm=1.0

Optimizer Step
AdamW
Update 14.8M params
zero_grad()

★ DETACH MEMORIES ★
compressed_memories = [
m.detach() for m in memories
]
Breaks gradient flow
Preserves activations

More batches?

Yes
(memory
persists)

No

End Epoch
Save checkpoint