Chatbot

- **Video:** Tasks with Long Sequences
  2 min
- **Reading:** Tasks with Long Sequences
  10 min
- **Reading:** Optional AI Storytelling
  15 min
- **Video:** Transformer Complexity
  3 min
- **Reading:** Transformer Complexity
  10 min
- **Video:** LSH Attention
  4 min
- **Reading:** LSH Attention
  10 min
- **Reading:** Optional KNN & LSH Review
  20 min
- **Lab:** Ungraded Lab: Reformer LSH
  1h
- **Video:** Motivation for Reversible Layers: Memory!
  2 min
- **Reading:** Motivation for Reversible Layers: Memory!
  10 min
- **Video:** Reversible Residual Layers
  5 min
- **Reading:** Reversible Residual Layers
  10 min
- **Lab:** Ungraded Lab: Revnet
  1h
- **Video:** Reformer
  2 min
- **Reading:** Reformer
  10 min
- **Reading:** Optional Transformers beyond NLP
  20 min
- **Reading:** Acknowledgments
  10 min

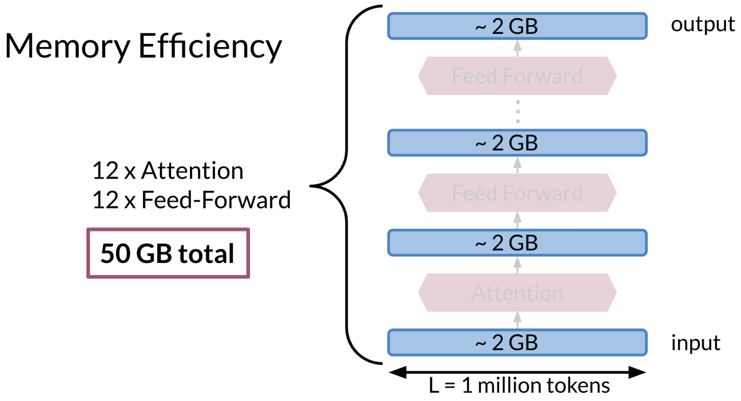
Heroes of NLP: Quoc Le

Assignment

Course Resources

Motivation for Reversible Layers: Memory!

Every time you run a forward propagation, you need to compute the back propagation to update the weights. The biggest issue with doing this is that you have to store the weights to be able to compute the back-prop. With these very large models, that could be a lot of memory.



Memory Efficiency

12 x Attention
12 x Feed-Forward

50 GB total

For example in the model above it requires 2GB to compute the Attention and 2GB for the feed forward. You have 12 layers for attention and 12 layers for the feedforward. That is equal to 12 * 2 + 12*2 + 2 (for the input) = 50 GB. That is a lot of memory. In the next video you will learn how to solve such problems.

Mark as completed