ALBERT-LSH Structure

Add & Norm

Feed Foward

*12 Layers

Add & Norm

Locality Sensitive Hashing Attention

Share Parameters

Sentence Order Loss

Masking Loss

ALBERT

Masking

Seq A   Seq B   Seq C   ...

Pre-training

Task B

Task A

Out A   Out B   Out C

ALBERT w.

Locality Sensitive Hashing

Seq A   Seq B   Seq C   ...

Fine-tuning