­

ALBERT-LSH Structure

Fine-tuning

Task B

Task A

…

Out C

Out B

Out A

ALBERT w.

Locality Sensitive Hashing

Seq C

Seq B

Seq A

Pre-training

Masking

Loss

Sentence

Order Loss

Seq C

Seq B

Seq A

…

ALBERT

Masking

\*12 Layers

Share

Parameters

­­

Add & Norm

**Locality Sensitive Hashing**

**Attention**

Add & Norm

Feed

Foward