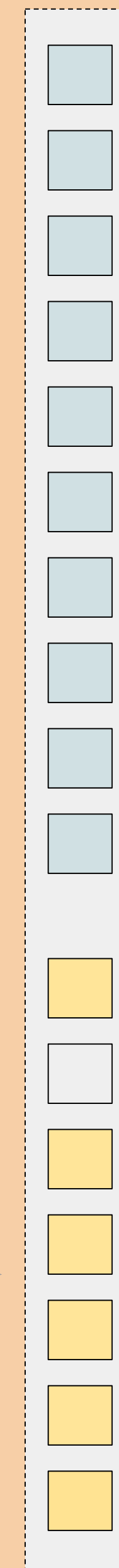




Vision Tower
SigLIP2-B16

Pixel Shuffling

Linear Projection



cat

"The [MASK] sat
on the mat."

Tokenizer

ModernVBERT Architecture