

How much knowledge  
is useful?

$$H = m_p \circ ((\mathbf{W} + s \cdot m_o \circ W_B(m_r \circ W_A \circ m_i))X + \mathbf{b})$$

Freeze

Pre-trained knowledge

Tuning

Task-specific knowledge