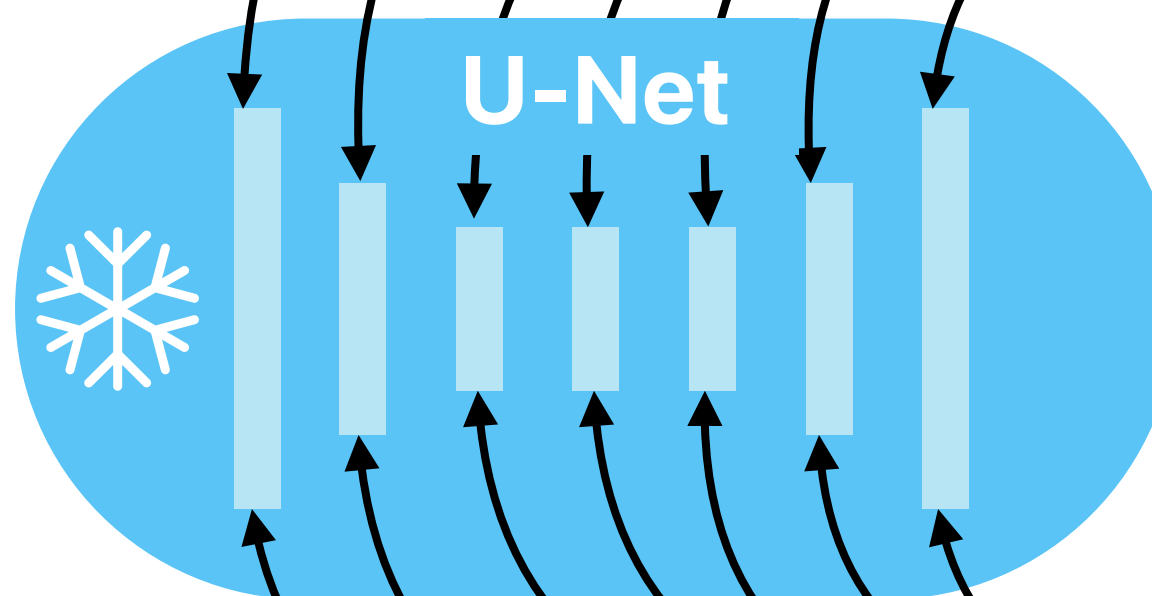


A serious man →

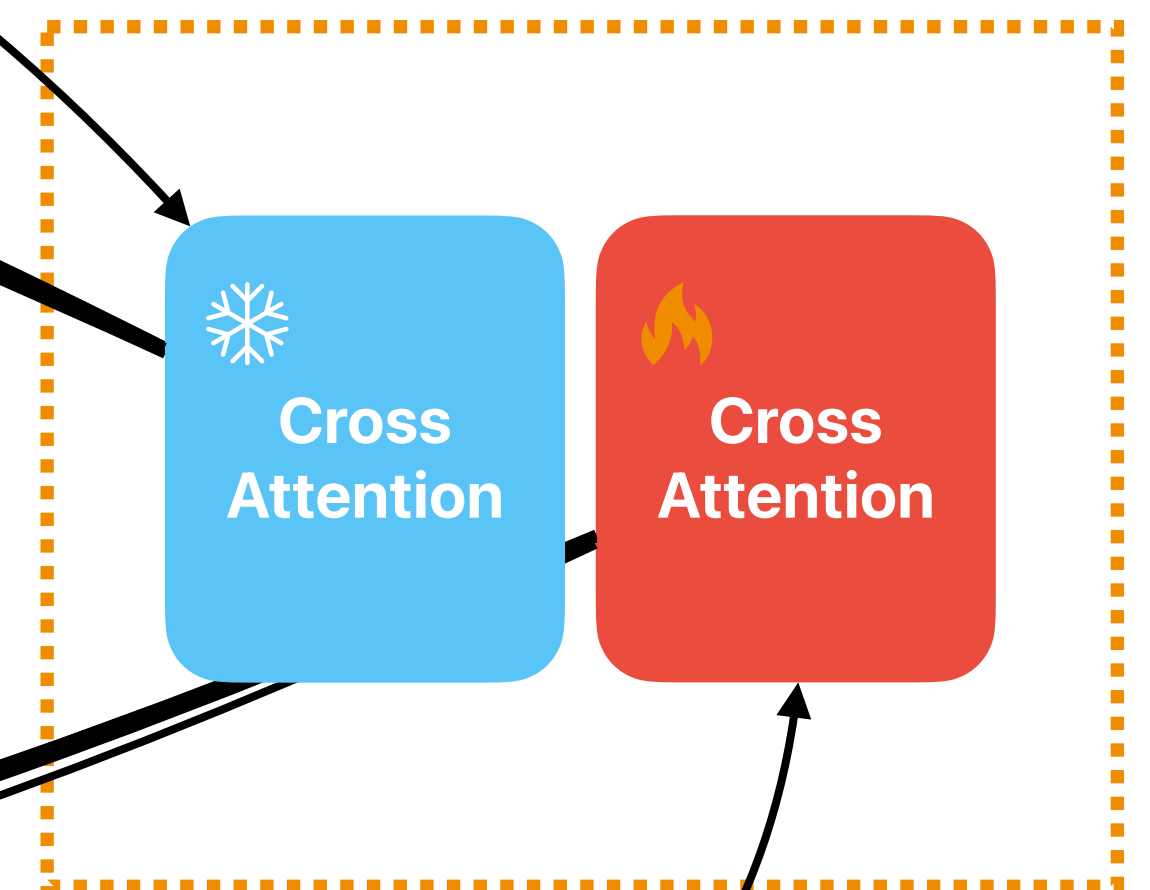


Text Features

A blue rounded rectangle representing the output of the text encoder.



Decoupled Cross-Attention



Images Features

A red rounded rectangle representing the output of the Lin and LN modules.

