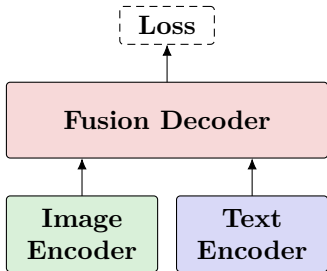


Dual-Encoder (Like CLIP)



Fusion (Like CoCa)