

Mapping Visual Features



$$mult_t = [W_m^V V_t \oplus W_m^L L_t \oplus W_h h_{t-1}^\delta]$$



Mapping Language Features