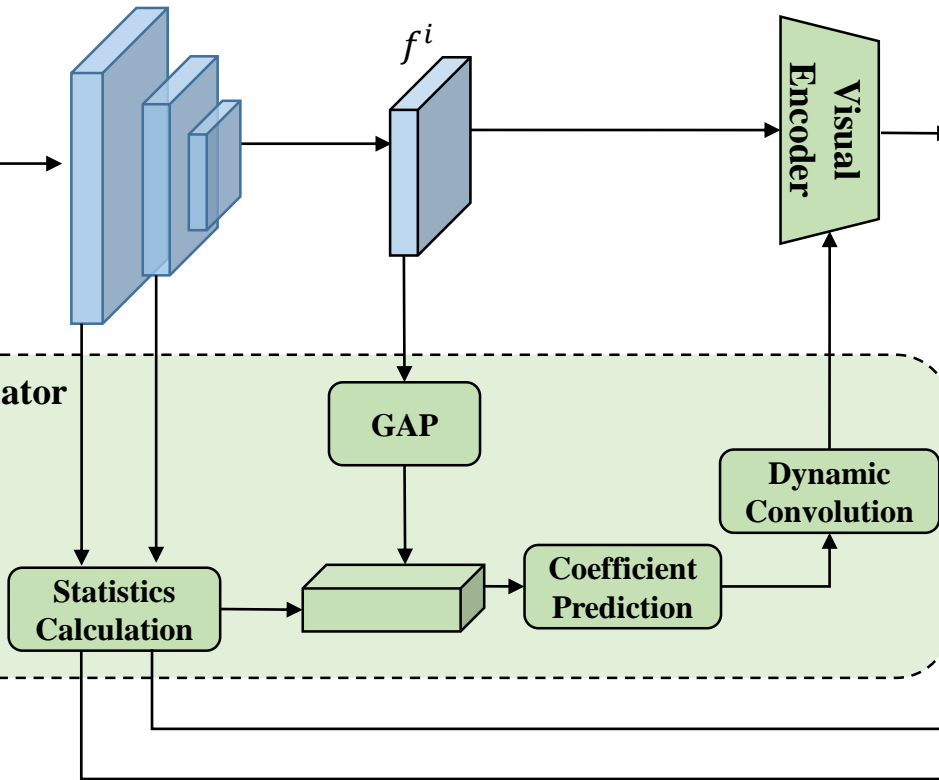




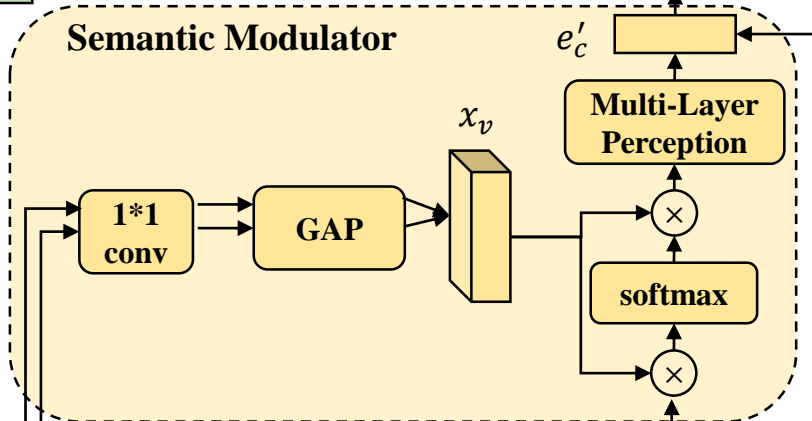
Visual Modulator



Visual Encoder

e_v^i

Semantic Modulator



Object-Conditioned Network

e_c

Glove

attributes

objects

Dot-Product Similarity

predictions