

Multimodal Encoder-Decoder Architecture with Cross-Modal Attention

