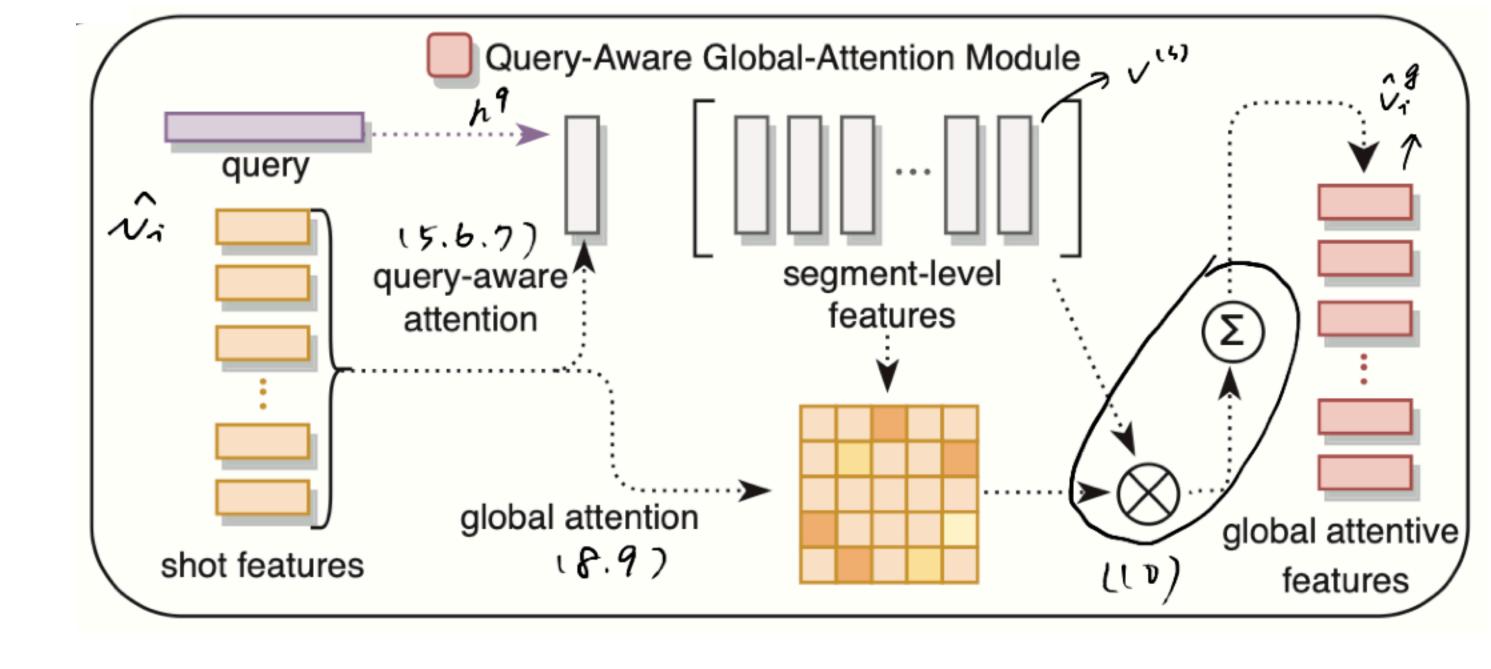
Proposed Method

Query global-attention module

- Compute the query-aware global -attentive representation for each shot.
- Given visual feature \hat{v}_i & all segment-level visual representation $(v_1^{(s)}, v_2^{(s)}, \dots, v_m^{(s)})$
 - *m*: number of video segments



(8)
$$e_j^g = v^T \tanh(W_1^g \hat{v}_i + W_2^g v_j^{(s)} + b)$$

(9)
$$r_j^g = \frac{\exp(e_j^g)}{\sum_{k=0}^m \exp(e_k^g)}$$

(10)Global-attentive representation for i-shot:

$$\hat{v}_j^g = \sum_{j=0}^m r_j^g v_j^s$$

Proposed Method

Deconvolutional Layer

Local Deconvolutional Self-Attention Use fully Propose several 1D convolutional block deconvolutional aggregated to decrease the layers to recover shot features shot features features the original number number on of video shots. temporal axis of video features. aggregated features shot features shot features query query Fully volutional Block _ayer nvolutional information aggregated features fusion shot features shot features

19ayer