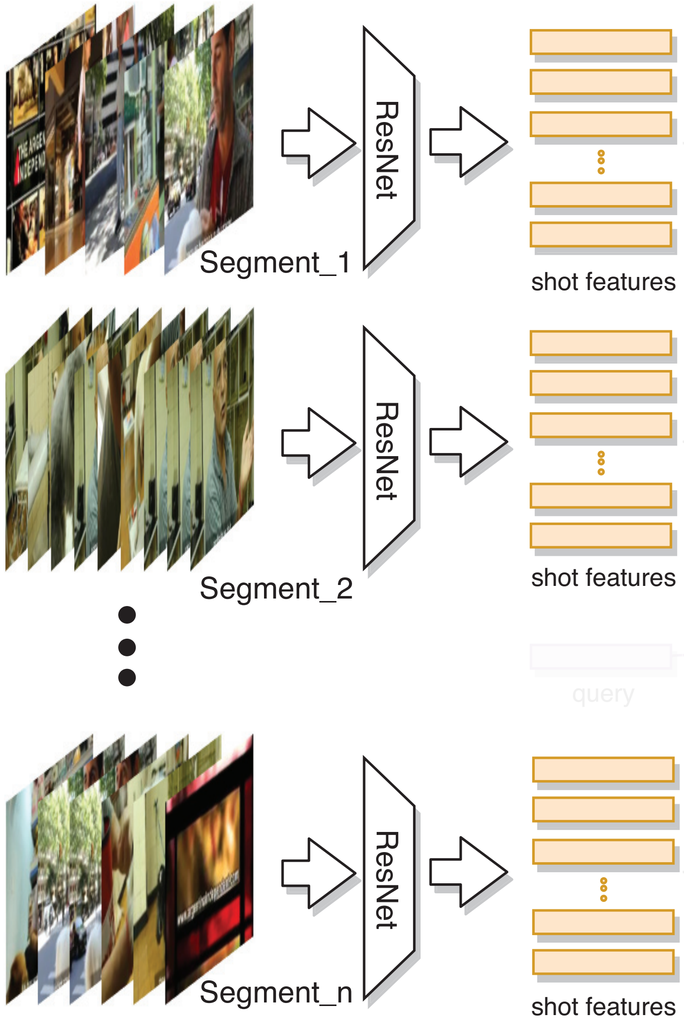
Proposed Method

Problem Formalization

- In the benchmark dataset, in where each query is consist of two concept (c_1, c_2)
  - Compute concept-related score for each shot
  - Then merge two kind of score as the query-related score
  - Finally, based on the score, can produce a diverse subset of video segments
    - Not only represent the origin video but related to the query
- **Input:** A long video v and a query q
- **Output:** A diverse subset of video shots remains original video info and related to query

Proposed Method

Feature Encoding Network



- Use the pertained ResNet to extract the visual feature of each shot after the segmentation process
- Denote the shot-level features in a specific video segment as $(v_1, v_2, \ldots, v_{s_k})$
  - s_k : the number of shots in the video segment
- Then use convolutional structure to further encode the visual feature