Introduction
Contribution

- Propose **Convolutional Hierarchical Attention Network**

  - Based on <u>convolution network</u> and <u>global-local attention</u> mechanism

  - Able to <u>generate video query-related summary</u> in parallel

- Present a **feature encoding network** to learn the features of each video shot

  - Employ <u>fully convolutional network</u> with <u>local self-attention and query-aware global attention</u> mechanism to obtain features with more semantic information

- Employ a **query-relevance computing module**

  - Takes the feature of video shot and <u>query as input and then calculate the similarity score</u>

Introduction
Convolutional Hierarchical Attention Network (CHAN)

- Consists of two parts:

  - A **feature encoding network** to learn features from each video shot **in parallel** from a local perspective and a global perspective.

  - **Query-relevance ranking module** to calculate the similarity score with respect to a query for each shot and then select video content related to the given query.