

# Introduction

## Differences between query-focused & generic video summarization

- Summary needs to take to the subjectivity of users into account
  - Different user queries may receive different video summaries
- Trained video summarizers can't meet all the users' preference
- Textual query will bring additional semantic information to the task

# Related Work

## Generic Video Summarization

- **Unsupervised:** take advantage of specific selection criteria to measure the importance or interestingness of video frames and then generate video summaries.
  - hand-crafted heuristics (2014), frame clustering (2016), GAN (2017)...
- **Supervised:** training data used in these methods contains raw videos and human-created ground-truth annotations. Combined with a large number of human annotations, models can capture the video content with more semantic information
  - Web image (2014), dppLSTM (2016), bi-LSTM (2016), RNN+LSTM (2018)...
- **Weakly supervised approaches:** readily available label, such as video categories, as additional information to improve model performance
  - 3D ConvNet (2017)