Experiments

Compared Models

- **SeqDPP (2014):** formulates video summarization as a subset selection problem and use sub-modular maximization to found summary. (dose not consider user queries)
- **SH-DPP (2016):** extension of SeqDPP, add a extra layer in the process of SeqDPP to judge whether a video shot is related to a given query.
- **QC-DPP (2017):** another extension of SeqDPP, introduces memory network to parameterize the kernel matrix.
- **TPAN (2018):** the three-player adversarial network, uses GAN to tackle with the task and introduce a random summary as an extra adversarial sample.

Experiments

Experimental Results

Table 1: Comparison results on the query-focused video summarization dataset in terms of Precision, Recall and F1-score.

	SeqDPP			SH-DPP			QC-DPP			TPAN			CHAN		
	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1
Vid1	53.43	29.81	36.59	50.56	29.64	35.67	49.86	53.38	48.68	49.66	50.91	48.74	54.73	46.57	**49.14**
Vid2	44.05	46.65	43.67	42.13	46.81	42.72	33.71	62.09	41.66	43.02	48.73	45.30	45.92	50.26	**46.53**
Vid3	49.25	17.44	25.26	51.92	29.24	36.51	55.16	62.40	56.47	58.73	56.49	56.51	59.75	64.53	**58.65**
Vid4	11.14	63.49	18.15	11.51	62.88	18.62	21.39	63.12	29.96	36.70	35.96	**33.64**	25.23	51.16	33.42
Avg.	39.47	39.35	30.92	39.03	42.14	33.38	40.03	60.25	44.19	47.03	48.02	46.05	46.40	53.13	**46.94**

- CHAN is outperforms the state-of-the-art approach (TPAN) by 1.9%
  - Specifically video 2 & 3, CHAN can have a better performance than TPAN (2.64%, 3.6%)
- The improvements of performance identify the effectiveness of our approaches to learn the relevance between the video shots and user's query.
  - The average running time of each video 134.4ms, shorter than TPAN 1.614s by 91.6%.