Automatic Video Summarization from Cricket Videos Using Deep Learning

PAPER ID: 85

Solayman Hossain Emon*, A.H.M Annur, Abir Hossain Xian, Kazi Mahia Sultana & Shoeb Mohammad Shahriar

Department of CSE,
Ahsanullah University of Science and Technology.

ICCIT 2020
THE 23rd INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY

Outline of the Presentation

- Research Problem
- Motivation
- Related Works
- Challenges Faced
- Methodology
- Dataset
- Experiment & Result
- Conclusion

Research Problem

✓ Automatically Creating Skims from the Cricket Videos

❑ How can we implement that?

- Using Deep Learning Based Method

Motivation

01

Viewing Experience

Summarized videos increase the viewing experience

02

Overload of Data

Overwhelming video volumes and variability

03

Application Potential

Wide range of applications requires automated summarization

04

Manual Analysis

Manual analysis demands a certain level of proficiency

Related Works

A Tabular Gist of the Reviewed Papers

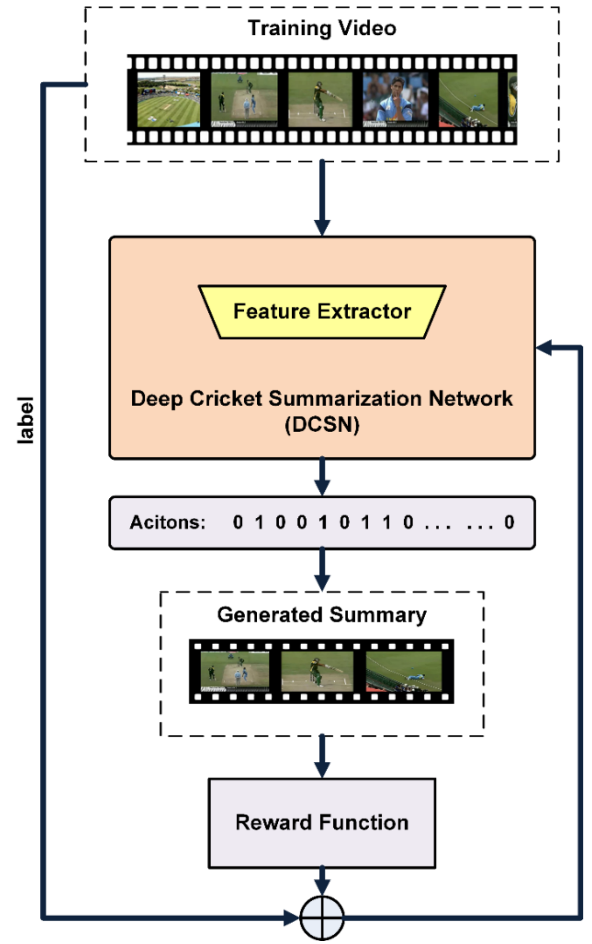
Contributors	Reference Work
Bhalla et al. 2019	Automatic Cricket Video Summarization.
Nasir et al. 2018	Event Detection and Summarization of Cricket Videos.
Zhang et al. 2016	Video Summarization with Long Short-Term Memory.
Zhou et al. 2018	Video Summarization with Deep Reinforcement Learning.
Mahasseni et al. 2017	Unsupervised Video Summarization with Adversarial LSTM Networks.
Panda et al. 2017	Weakly Supervised Summarization of Web Videos.

Challenges Faced

- Unavailability of Dedicated Dataset
- Learning Complex Patterns
- Lack of the Evaluation Process
- Subjective Nature

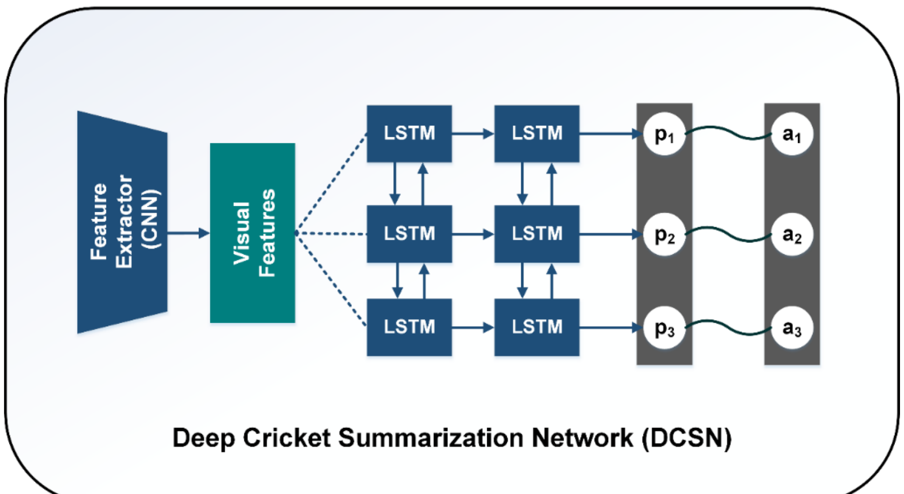
Methodology

- An encoder-decoder architecture for designing the summarization network.
- A Convolutional Neural Network (CNN) performs as encoder that extract the features.
- The Recurrent Neural Network (RNN) works as which capture the temporal information of the sequence.
- For training purpose, the formulation of Reinforcement Learning (RL) with Diversity-Representativeness (DR) reward function is used.



Deep Cricket Summarization Network (DCSN)

- Named the summarization network as Deep Cricket Summarization Network (DCSN)
- For developing the agent, a Long Short-Term Memory (LSTM) based bi-directional Recurrent Neural Network (RNN) with a Fully Connected (FC) layer is used.
- The network anticipates the frame-level probability (p_t).



Deep Cricket Summarization Network (DCSN)

$$\tilde{p}_t = sigmoid(W \times h_t)$$

$$\tilde{a}_t = Bernoulli(\tilde{p}_t)$$

Reward Function

A novel reward function (R_{dr}) composed of diversity reward (R_{div}) and representative reward (R_{rep}).

$$R_{dr} = \frac{1}{2} R_{\text{div}} + \frac{1}{2} R_{\text{rep}}$$

- ❑ **Diversity Reward**

$$R_{\text{div}} = \frac{1}{|K|(|K|-1)} \sum_{t \epsilon K} \sum_{\substack{\acute{t} \epsilon K \\ t \neq \acute{t}}} diss(x_t, x_{\acute{t}})$$

- ❑ **Representativeness Reward**

$$R_{\text{rep}} = \exp\left(-\frac{1}{T} \sum_{t=1}^{T} \min_{\acute{t} \epsilon K} (\|x_t - x_{\acute{t}}\|_2)\right)$$

Supervision Signals

$$\theta^* = \arg\max_{\theta} \sum_{j} \log\{P(s^{(j)} \subset S^{(j)}; L^{(j)})\}$$

Optimization

Objective Function, $J(\theta) = \mathcal{E}_{p\theta(\tilde{a}_1 : T)}[R_{dr}]$

Approximate Gradients Via REINFORCE:

$$\frac{\partial}{\partial \theta} J(\theta) \approx \frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} (R_{\alpha} - \rho) \nabla_{\theta} \log \pi_{\theta}(\tilde{a}_t | \tilde{h}_t)$$

Summary Generation

Kernel Temporal Segmentation (KTS) detects shot boundaries by detecting changes in visual features. The shot-level scores of k[th] shot is computed as:

$$S(F_k) = \frac{1}{|F_k|} \sum_{i \epsilon F_k} p_i$$

For selecting the shots of a summary with particular length ($ł_s$), we solve the following 0/1 knapsack optimization problem.

$$\arg\max \sum_k \mu_k S(F_k), \quad \sum_k \mu_k |F_k| \le l_s; \quad \mu_k \epsilon \{0, 1\}$$

Dataset

There is a limitation of availability of dedicated datasets in cricket match domain. For this remedy, we introduced a new cricket dataset namely CricSum.

- Extraction of Novel Features
- Normalization of Features
- Subsampling
- Detection of Shot Boundaries
- Ground Truth Preparation

Experiment & Result

- **Objective Analysis:** Analyzing the significant change in the score using Optical Character Recognition (OCR).

Accuracy (%) of Key Events Detection through Proposed Method

Method	Wickets	Boundaries (4)	Six (6)	Excitements
$DCSN_{sup}$	98.63%	91.87%	90.13%	94.32%

- **Subjective Analysis:** The automated summarized videos receives Mean Opinion Score (MOS) value 4 out of 5.

Experiment I

Evaluation by F-score (%) for Different Variants of Proposed Model

Dataset	Method	Feature Type	Video	F-score (%)	Average F-score (%)
CricSum	$DCSN_{sup}$	Deep	video_4	49.1	60.6
			video_7	69.8	
			video_3	67.3	
			video_2	56.3	
CricSum	$DCSN_{unsup}$	Deep	video_4	46.5	56.9
			video_7	63.6	
			video_3	67.3	
			video_2	50.4	

Experiment II

F-score (%) for Variants of Recurrent Neural Network (RNN) Units

Method	CricSum	
	LSTM	GRU
$DCSN_{sup}$	60.6	59.1
$DCSN_{unsup}$	56.9	52.8

Experiment III

Result (%) of Supervised Model According to Reward

Dataset	Method	Feature Type	RNN Units	Average F-score (%)
CricSum	$DCSN_{sup}$	Deep	LSTM	60.6
CricSum	$Div-DCSN_{sup}$	Deep	LSTM	59.10
CricSum	$Rep-DCSN_{sup}$	Deep	LSTM	59.35

Comparison Between Existing Works and Proposed Method

Accuracy (%) of Key Events Detection Between Proposed & Existing Methods

Method	Wickets	Boundaries (4)	Six (6)	Excitements
Bhalla et al. [19]	92.45%	86.74%	89.12%	-
Nasir et al. [16]	98.04%	90.78%	92.97%	-
Proposed Method	98.63%	91.87%	90.13%	94.32%

Conclusion

- Achievements of the Proposed Approach
- Limitations
- Future Scopes

Question Answer Session

Thanks for your patience