Video Summarization with a Dual Attention Capsule Network

Hao Fu\textsuperscript{1}, Hongxing Wang\textsuperscript{1}, Jianyu Yang\textsuperscript{2}
\textsuperscript{1}Chongqing University, China
\textsuperscript{2}Soochow University, China

\begin{itemize}
\item Problem
\end{itemize}
Generating a compact and non-redundant summary for a given video without missing significant information.

\begin{itemize}
\item Unrepresentative
\item Redundant
\item Good summary
\end{itemize}

\begin{itemize}
\item Contributions at a Glance
\end{itemize}
\begin{itemize}
\item We propose a novel dual attention capsule network model, which can effectively incorporate the short- and long-term temporal dependencies among video frames for summarization.
\item Our proposed video summarization is parallelizable, which can easily handle longer-term dependencies among video frames than the RNN/LSTM-based approaches.
\item Experimental results show that our proposed method owns stronger learning ability, and is competitive with existing state-of-the-art methods.
\end{itemize}

\begin{itemize}
\item Method
\end{itemize}

\textbf{Feature Extraction}
- Extractor: GoogLeNet
- Divided frames into M clips
- Each clip is viewed as a local block

\textbf{Dual Attention Feature Refinement}
- Self-attention mechanism
- Learn the short- and long-term dependencies within and between clips

\textbf{Two-Stream Capsule Network Learning}
- Later fusion mechanism
- Scalar $\rightarrow$ Vector
- Margin loss

\begin{itemize}
\item 
\end{itemize}

\begin{itemize}
\item Experimental Results
\end{itemize}

\begin{itemize}
\item Ablation study of our method
\item $\tau$ and $\rho$ results compared with SOTA
\item F-score(\%) results compared with SOTA
\end{itemize}

\begin{itemize}
\item References
\end{itemize}