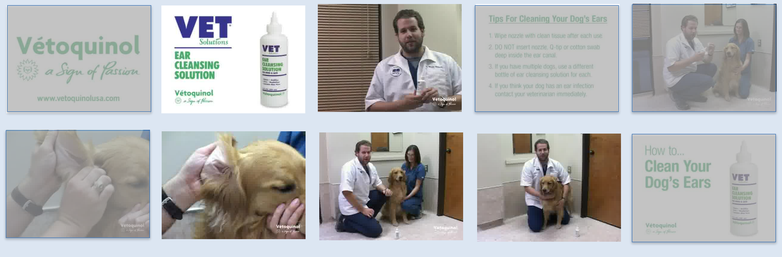
Unsupervised Video Summarization with Adversarial LSTM Networks

Behrooz Mahasseni, Michael Lam and Sinisa Todorovic

Problem: Video Summarization

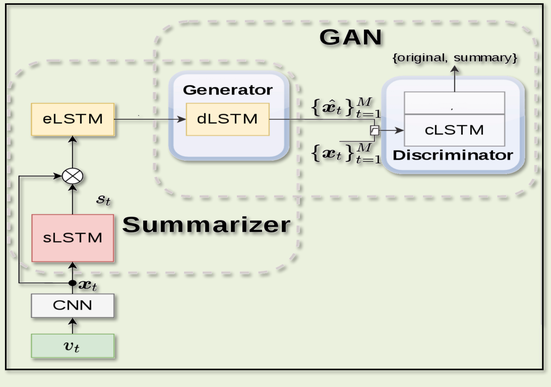
- Goal : Select a sparse set of key frames



Key Idea: Adversarial Training

- The summary should have the same content as the original → GAN
- Sparse selection of frames should represent well the original → VAE
- Contribution:
  - ❖ Unsupervised Summarization with VAE + GAN

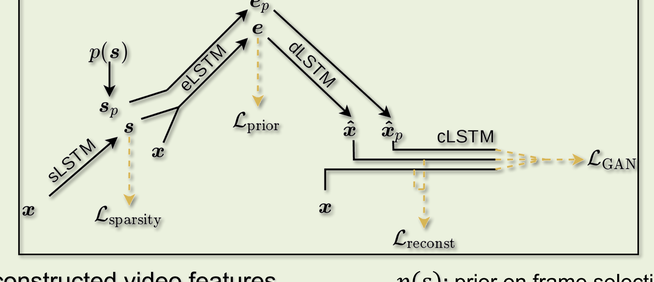
Adversarial Video Summarizer



$\hat{x}$: reconstructed video features
x: original video features

s_t: frame selection
v_t: frame at time t.

Regularized Training



$\hat{x}$: reconstructed video features.
x: original video features
$\hat{x}_p$: reconstructed from uniform sampling

$p(s)$: prior on frame selection.
e: encoder output.

Accuracy in [%] on SumMe & TVSum

Method	SumMe	TVSum
Avila et al. '11	33.7	-
Li et al. '10	26.6	-
Lhoslaet al. '13	-	36.0
Song et al. '15	26.6	50
Zhao et al. '14	-	46.0
Ours	39.1	51.7

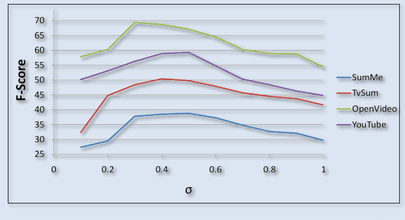
(unsupervised)

Method	SumMe	TVSum
Avila et al. '11	33.7	-
Li et al. '10	26.6	-
Lhoslaet al. '13	-	36.0
Song et al. '15	26.6	50
Zhao et al. '14	-	46.0
Ours	39.1	51.7

(augmented training set)

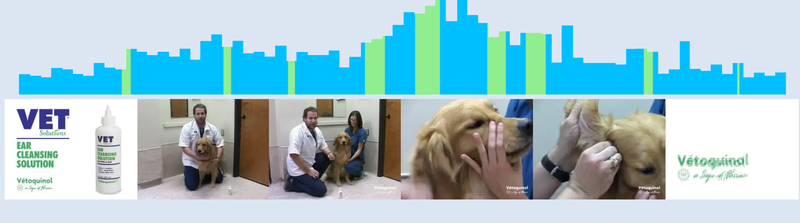
Method	SumMe	TVSum
Gygli et al. '15	39.7	-
Zhang et al. '16	40.9	-
Gygli et al. '14	39.3	-
Zhang et al. '16	38.6	54.7
Ours	39.1	51.7

(supervised)



(F-score for variations of σ)

(An Example Video Summarization)



Acknowledgments: DARPA XAI & NSF RI1302700.