A Framework towards Domain Specific Video Summarization

Vishal Kaushal, Sandeep Subramanian, Suraj Kothawade, Rishabh Iyer, Ganesh Ramakrishnan

IEEE 2019 Winter Conference on Applications of Computer Vision

MOTIVATION

- **Unprecedented growth in video data**
  - Makes it challenging to store and consume
  - Video summarization attempts to address these challenges, but still an unsolved problem
- **What constitutes a good summary varies from domain to domain**
  - Sports video - ‘importance’ is more important than ‘diversity’
  - Surveillance video - ‘diversity’ is more important than ‘importance’
  - Further, what is ‘important’ for soccer videos is different from what is ‘important’ for birthday videos
- **A good video summary of a domain should have both characteristics!**

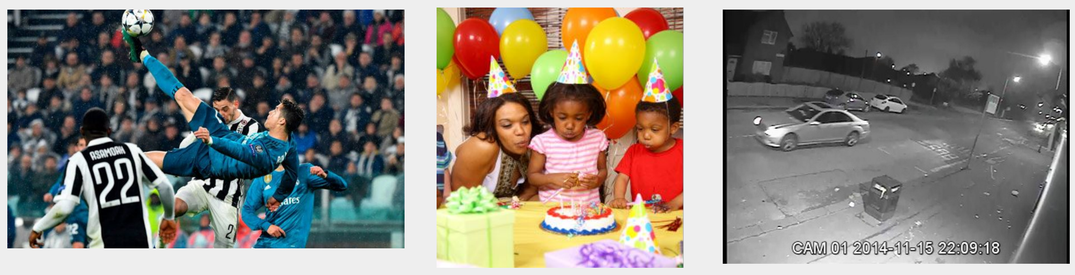
OUR CONTRIBUTIONS

- **Joint problem** of learning domain specific importance of segments as well as the desired summary characteristic for that domain
- An effective rating mechanism to serve as **indirect ground truth**
- A **novel evaluation measure**, more naturally suited in assessing the quality of video summary for the task at hand than F1 like measures
- A **gold standard dataset** for domain specific video summarization, first known dataset of long videos

METHODOLOGY

- **Weighted mixture** of modular and submodular terms
  - **Modular** terms: capture the domain specific importance of snippets
  - **Submodular** terms like Set Cover, Facility Location etc.: impart certain desired characteristics to the summary
- Learn weighted mixture using **max margin learning framework**
  - Different weights learnt for different domain
- For any given test video of that domain, the **weighted mixture is then maximized** to produce the desired summary video

GOOD SUMMARY?



Formulation

$$y^* = \underset{y \subseteq Y_v, |y| \leq k}{\operatorname{argmax}} \; o(x_v, y)$$

$$o(x_v, y) = w^T f(x_v, y)$$

$$\min_{w \geq 0} \frac{1}{N} \sum_{n=1}^{N} L_n(w) + \frac{\lambda_1}{2}||w_1||^2 + \frac{\lambda_2}{2}||w_2||^2$$

$$L_n(w) = \max_{y \subseteq Y_v^n} (w^T f(x_v^n, y) + l_n(y)) - w^T f(x_v^n, y_{gt}^n)$$

Evaluation Measure

$$S_V(y) = \sum_{x_i \in X_P} |y \cap x_i| * (1 + \frac{|y \cap x_i|}{|x_i|}) * e^{\alpha * rating(x_i)} + \sum_{x_i \in X_R} \min(|y \cap x_i|, \beta) * (1 + \frac{\min(|y \cap x_i|, \beta)}{\min(|x_i|, \beta)}) * e^{\alpha * rating(x_i)} - \sum_{x_i \in X_N} |y \cap x_i| * k$$

Positively Rated: Reward

Repetitive: Saturate

Negatively Rated: Penalize

RESULTS & CONCULSIONS

- **Full mixture** works best
- Models trained on one **domain** do not perform well on another
- **Multiple ground truths** help
- **Strong correlation** between components based on learnt weights and components with highest individual score when optimized
- **Intuitive relationship** of individual components with domains

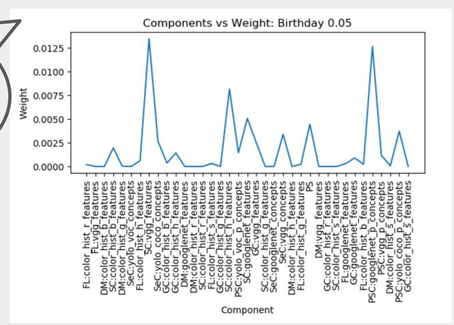
Domain	Method	ScoreLoss
Birthday	All Modular	0.7234
	All Submodular	0.7307
	Full	**0.6625**
	Random	0.7378
	Uniform	0.7569
	Submodular	0.7432
EntryExit	All Modular	0.5967
	All Submodular	0.6306
	Full	**0.5884**
	Random	0.7706
	Uniform	0.7785
	Submodular	0.6306
Cricket	All Modular	0.8140
	All Submodular	0.8275
	Full	**0.7733**
	Random	0.8911
	Uniform	0.8979
	Submodular	0.8275
Office	All Modular	0.3871
	All Submodular	0.4783
	Full	**0.3696**
	Random	0.5743
	Uniform	0.5399
	Submodular	0.5590
Soccer	All Modular	0.8849
	All Submodular	0.7645
	Full	**0.6533**
	Random	0.9217
	Uniform	0.8747
	Submodular	0.9152

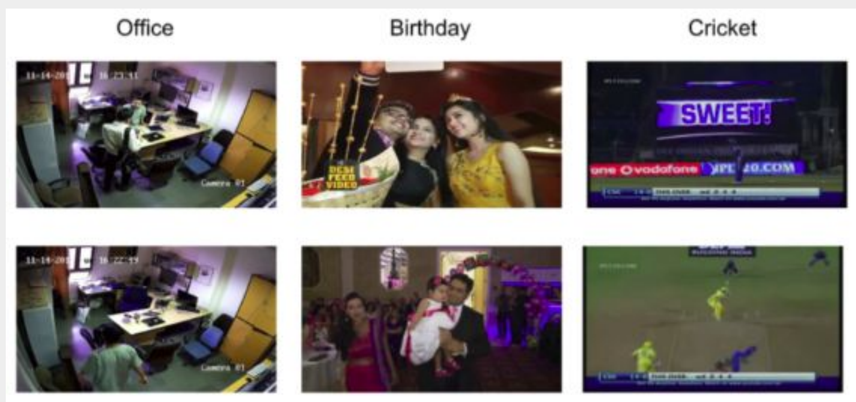
Birthday	Mod:vgg_features SC:color_hist_r_features PSC:yolo_coco_p_concepts SeC:yolo_coco_concepts SC:color_hist_b_features	PSC:googlenet_p_concepts SC:googlenet_features PS GC:vgg_features FL:vgg_concepts
Cricket	Mod:googlenet_features SC:vgg_features PSC:googlenet_p_concepts FL:googlenet_features Mod:googlenet_features	GC:googlenet_features FL:color_hist_g_features PSC:vgg_p_concepts GC:vgg_features PS
Office (Surveillance)	Mod:yolo_coco_p_concepts PSC:color_hist_b_features GC:color_hist_r_features DM:yolo_coco_concepts DM:yolo_coco_features	Mod:vgg_concepts Mod:vgg_features DM:vgg_features DM:color_hist_b_features DM:color_hist_r_features

Object Features
Scene Features
Object Detectors (People etc.)
Color Features
Diversity Model
Representation Model
Coverage Model
Importance Model

Different terms relevant for **different domains**

Full mixture performs the best, as hypothesized





Qualitative verification of video summaries produced