Temporal Saliency Adaption in Egocentric Videos

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Barcelona Supercomputing Center Centro Nacional de Supercomputación


Panagiotis Linardos

Eva Mohedano

Monica Cherto

Cathal Gurrin

Xavier Giro-i-Nieto

DCU DUBLIN CITY UNIVERSITY

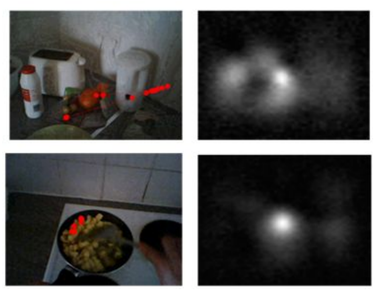
Insight

EPIC @ ECCV18

The Third International Workshop on Egocentric Perception, Interaction and Computing

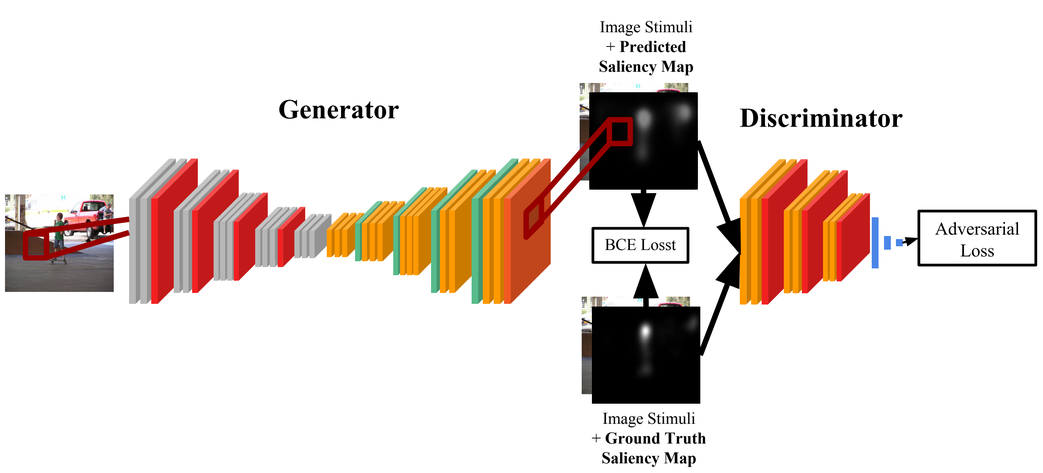
ECCV 2018 European Conference on Computer Vision 8 – 14 September 2018 | Munich, Germany

Motivation

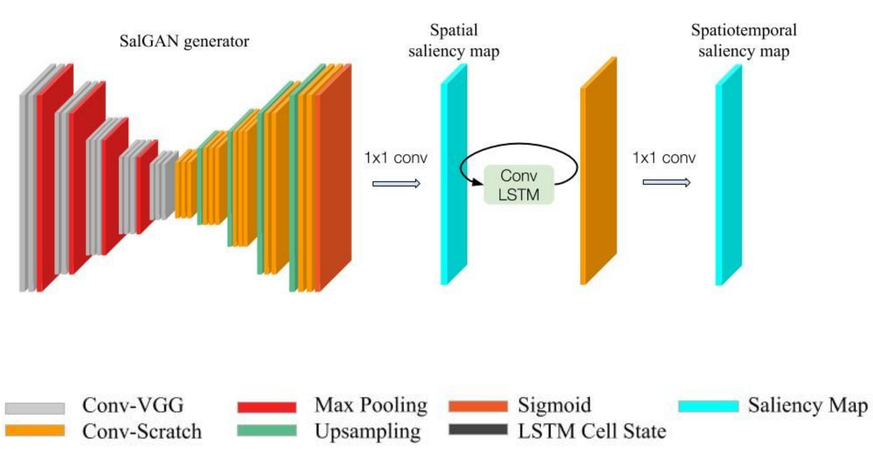


- Saliency prediction refers to the task of estimating which regions of an image have a higher probability of being observed by a viewer.
- It has been shown that this information can be helpful to improve tasks such as activity recognition or object detection.
- In this work, we focus in the task of **predicting saliency in egocentric videos**. For that, we extend a state of the art saliency model to the temporal domain.

SalGAN [1]



SalGAN [1] + convLSTM



Saliency maps from both configurations available for [2]:

EPIC KITCHENS

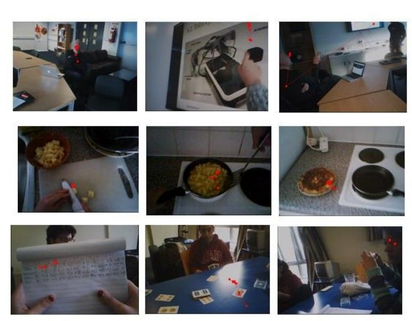
EgoMon Gaze & Video Dataset

New egocentric dataset

- 7 sequences: 4 free-viweing + 3 task-oriented
- Average length of 30 minutes
- 3 different users wearing Tobii glasses

Free-viewing activities

Task-oriented activities



...and also images from Narrative clip for one sequence..

Results

Performance on the DHF1K dataset

	AUC-J	sAUC	NSS	CC	SIM
DHF1K SoA	0.885	0.553	2.259	**0.415**	0.311
SalGAN	**0.930**	**0.834**	**2.468**	0.372	**0.264**
+ conv	0.743	0.723	2.208	0.303	0.261
+ convLST	0.744	0.722	2.246	0.302	0.260

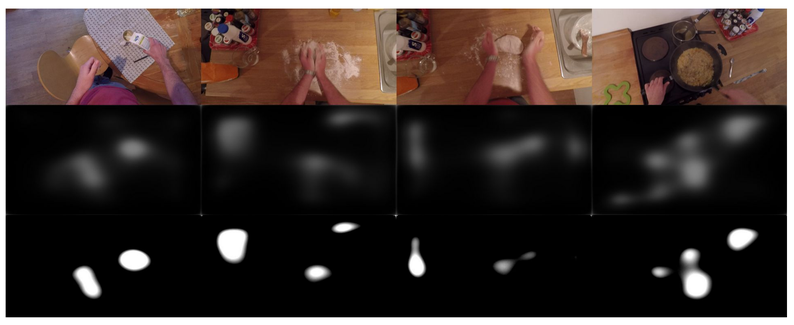
Performance on EgoMon tasks (NSS metric)

	Free-Viewing	Task-oriented	Total
SalGAN	**2.652**	1.313	**2.079**
+ conv	0.805	1.694	1.249
+ convLST	0.904	**1.705**	1.247

Video frames

SalGAN

+ convLSTM



Results indicate that SalGAN represents a strong baseline for video-saliency prediction. However, our strategy of including time information based directly on saliency predictions ignores any semantic information of the scene, so we will extend the work by working with lower SalGAN layers.

PyTorch

Model and source code: http://bit.ly/videosalgan

References:
[1] Junting Pan et al. "Salgan: Visual saliency prediction with generative adversarial networks." arXiv preprint (2017).
[2] Dima Damen et al. "Scaling Egocentric Vision: The EPIC-KITCHENS Dataset." ECCV 2018.

Science Foundation Ireland sfi For what's next

Generalitat de Catalunya

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

NVIDIA