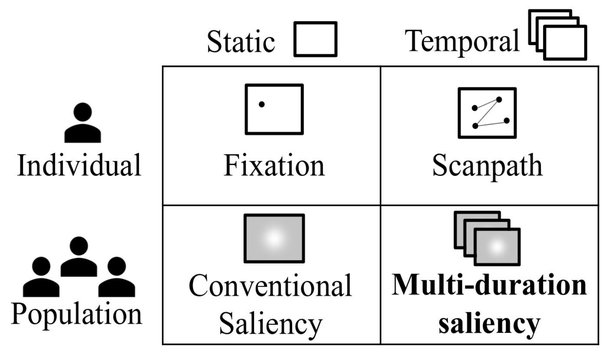
How many glances? Modeling multi-duration saliency

Camilo Fosco*, Anelise Newman*, Patr Sukhum, Yun Bin Zhang, Aude Oliva, and Zoya Bylinskii

What is multi-duration saliency?

A **rich, robust** representation of attention over time

- Has the generalizability of a population-level metric
- Contains temporal information
- Easy to collect and crowdsource

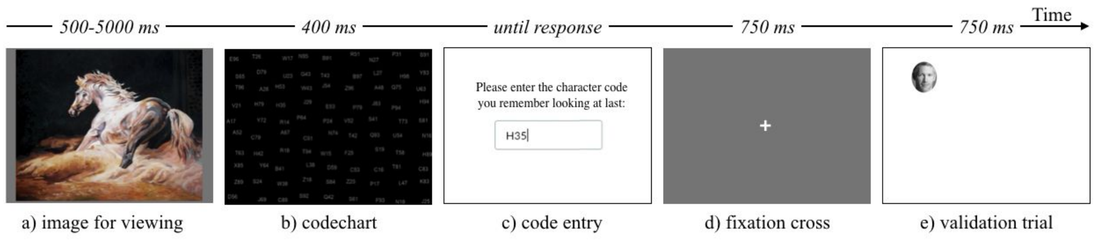


Data collection

We use the **CodeCharts** interface to collect gaze fixations at precise viewing durations.

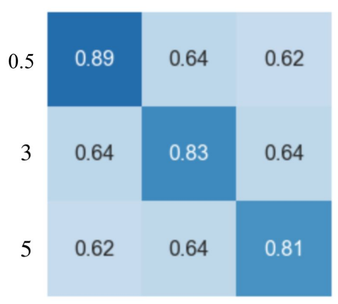
We collect saliency data at 0.5, 3, and 5 seconds.

Try out the interface!



CodeCharts1k

Introducing **CodeCharts1k**, the first multiduration saliency dataset.



Saliency varies by viewing duration. Split-half consistency is significantly higher within than across durations (measured using Pearson's Correlation Coefficient).

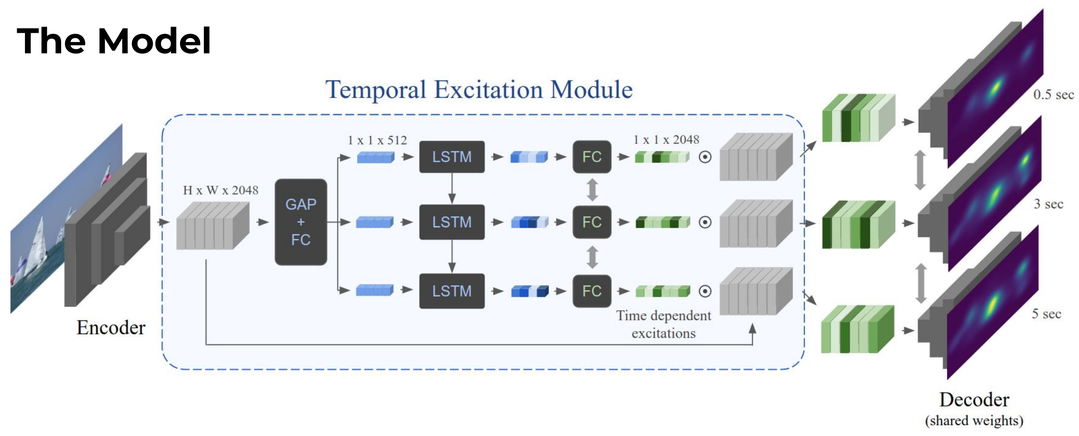
Temporal patterns in face saliency

The **boomerang** pattern: attention moves away from faces at 3s and back to faces at 5s.

The **decreasing** pattern: attention on faces decreases at 3s and again at 5s.

Modeling multi-duration saliency

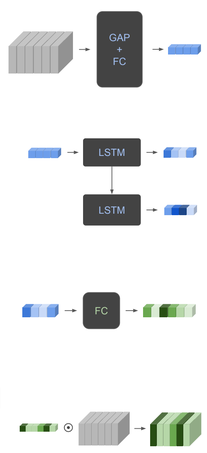
The Model



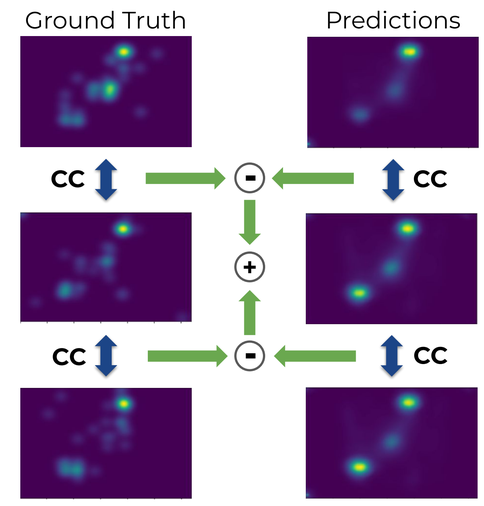
The Temporal Excitation Module

Produces multiple saliency maps with fewer params than comparable models.

- Compresses feature maps
- Applies iterative alterations with an LSTM
- Maps back into the original dimensionality
- Excites each channel of the original feature maps



The CCM Loss



We calculate the Correlation Coefficient (**CC**) for pairs of ground truth maps and pairs of predicted maps, and we minimize their difference.

Results

CodeCharts1k

Model	NSS	CC	KL
SAMx3	2.708	0.734	0.483
SAM-MD	2.818	0.761	0.446
MD-SEM (Ours)	**2.875**	**0.772**	**0.425**

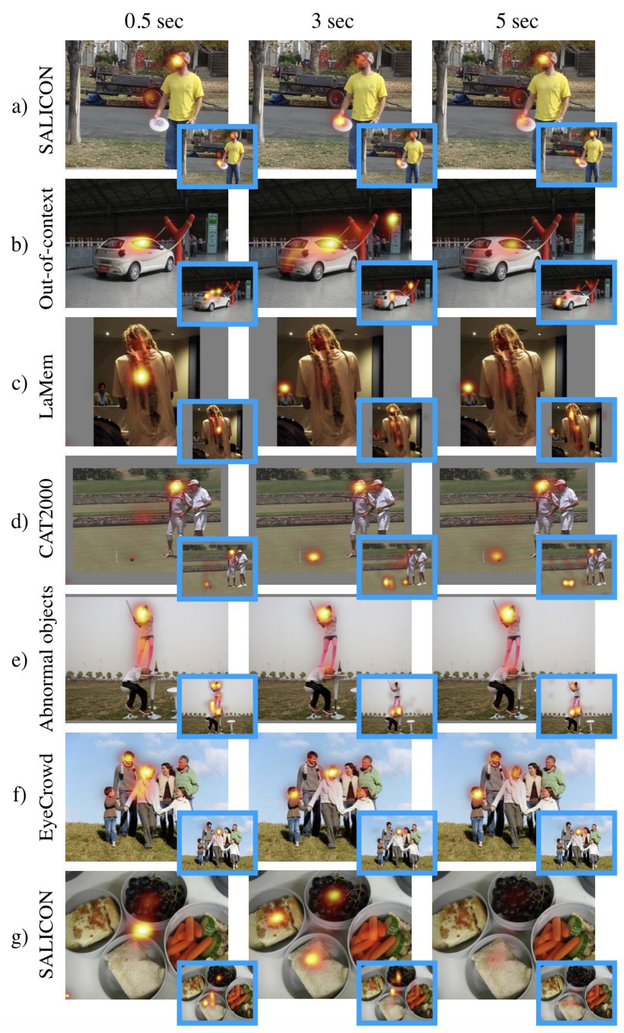
SALICON

Model	NSS	CC	KL
SAM-res	1.990	**0.899**	0.610
EML-Net	2.050	0.886	**0.520**
MD-SEM (Ours)	**2.058**	0.868	0.568

SALICON-MD*

Model	NSS	CC	Params
SAMx3	2.020	0.803	210M
SAM-MD	2.057	0.792	70M
MD-SEM (Ours)	**2.061**	**0.811**	**30M**

*SALICON split into different durations for pre-training



Applications

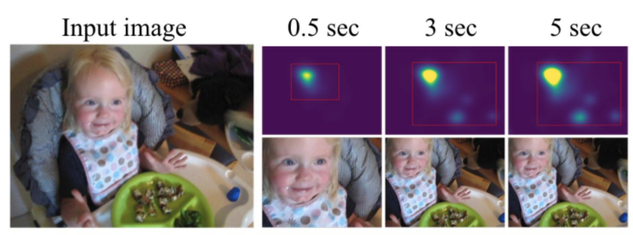
Captioning

Focus a captioning module on content that is salient at different durations

Rendering

Prioritize content to render based on order in which it is salient

Cropping



Generate image thumbnails/ summaries tailored to a certain duration