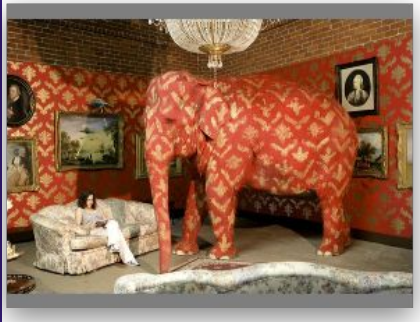
Multi-duration Saliency

Original Image



0.5 seconds

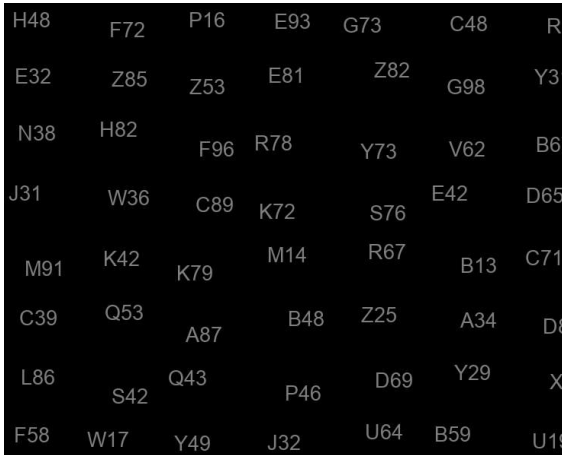
3 seconds

5 seconds

The CodeCharts Interface

500-5000 ms

400 ms



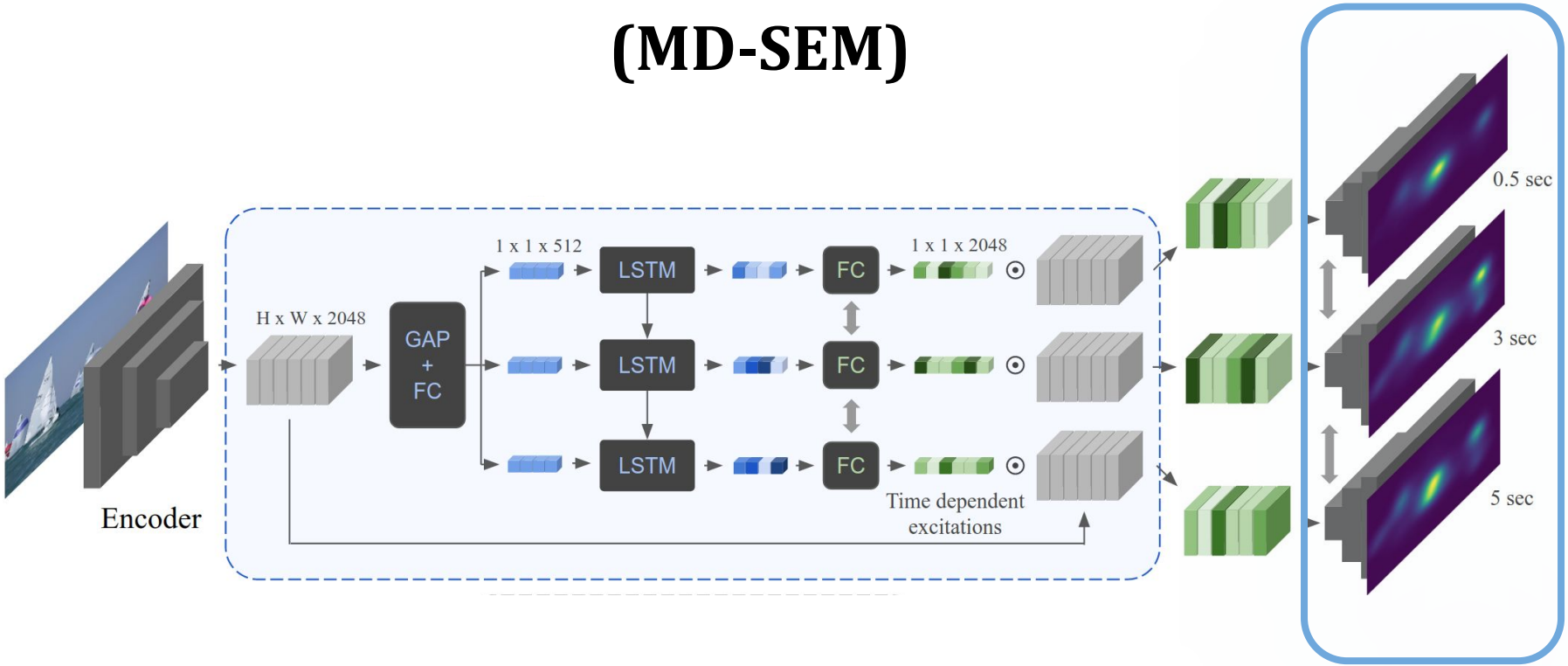
Please enter the character code you remember looking at last:

F96

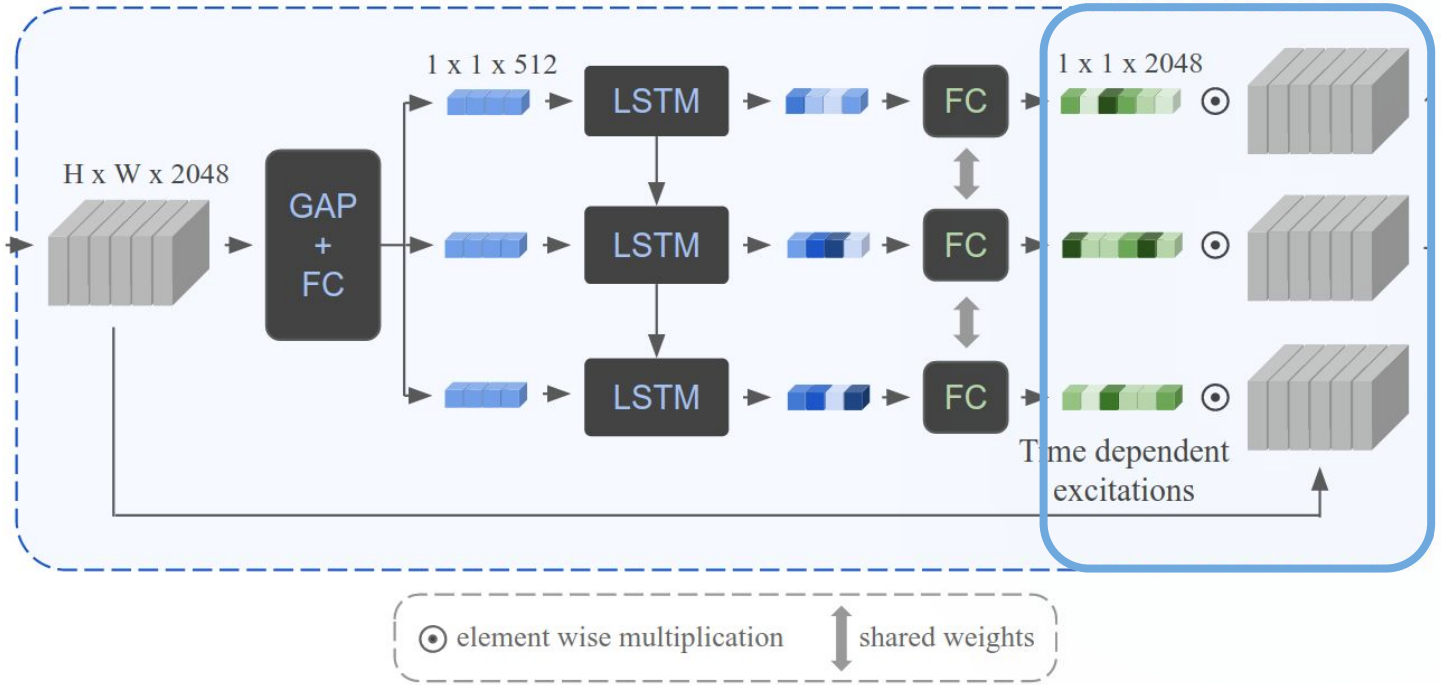
CodeCharts1k

	0.5 seconds	3 seconds	5 seconds
CAT2000 (Actions)			
Out of Context			
SALICON			

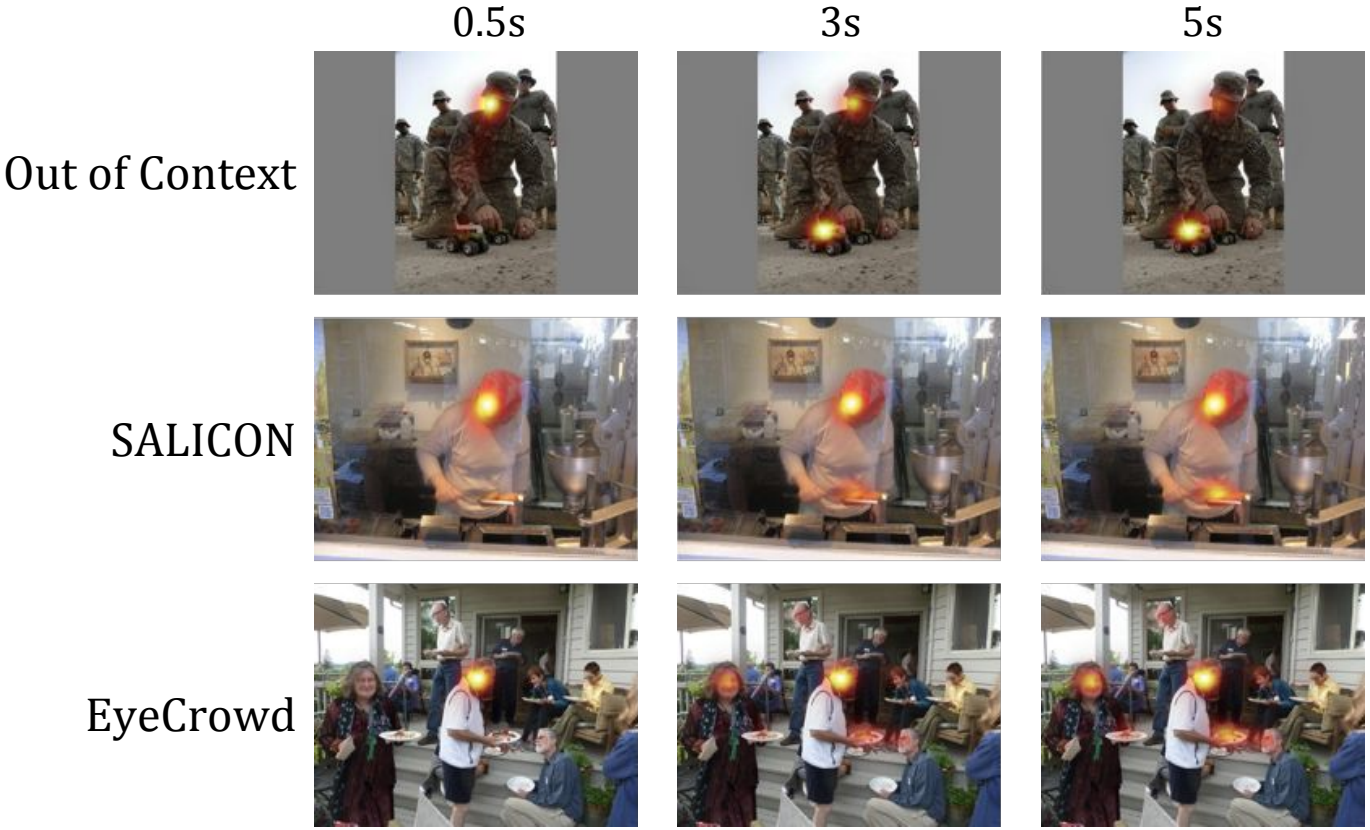
Multi-Duration Saliency Excited Model (MD-SEM)



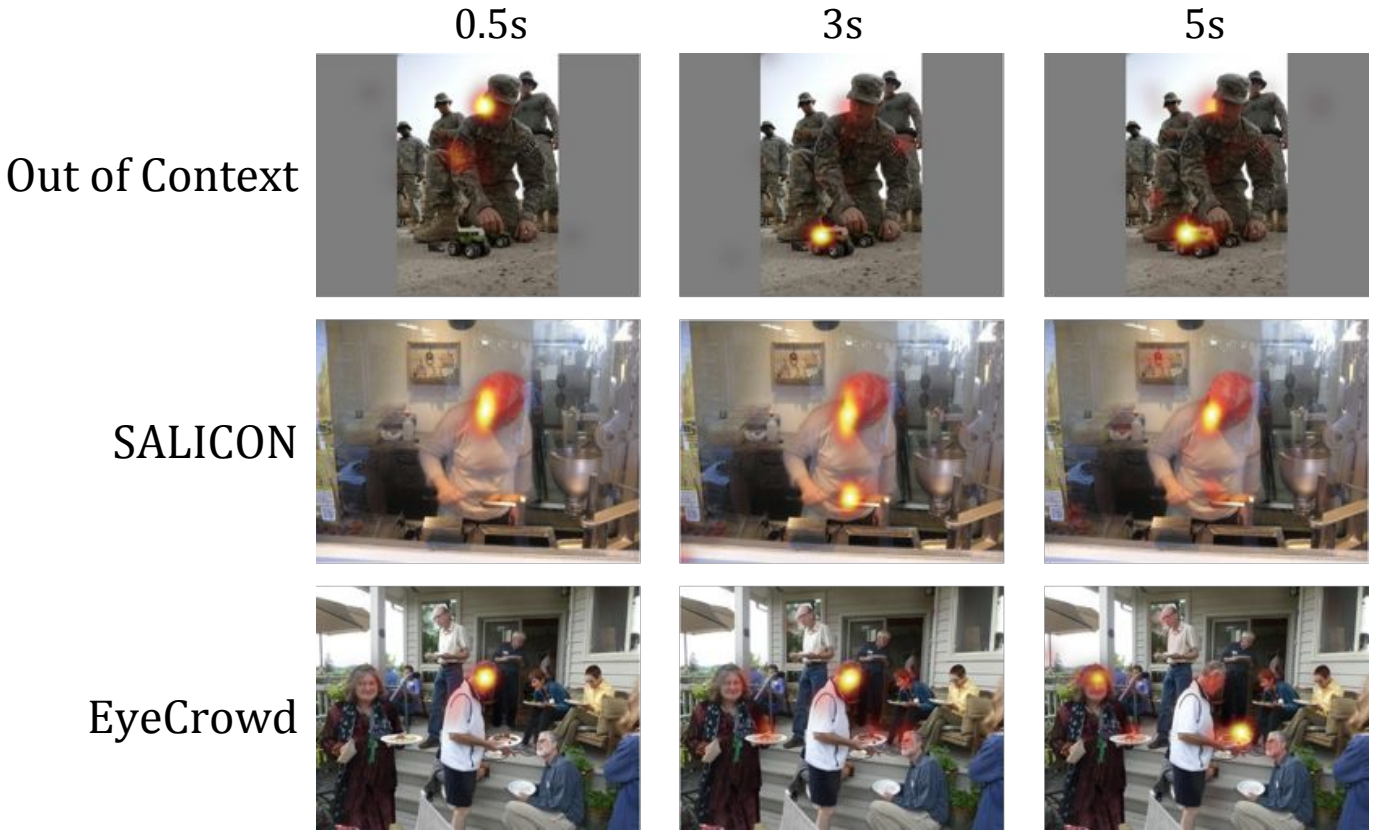
Temporal Excitation Module



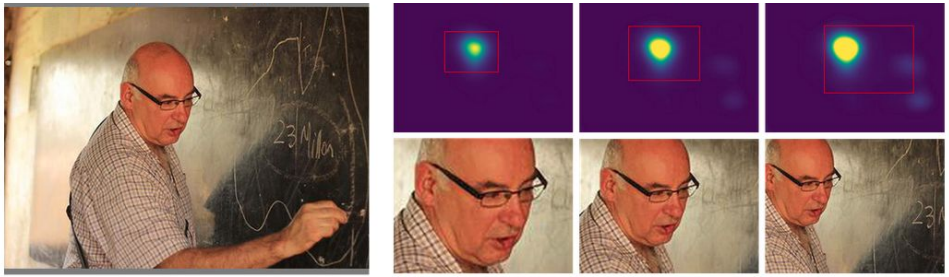
MD-SEM Predictions



CodeCharts1k Ground Truth



Applications



Cropping

Rendering

"two men sitting at a table with a vase of flowers"

"two men in suits standing in front of a table with a"

"two men in suits standing next to a table with flowers"

Captioning

How much time do you have? Modeling multi-duration saliency

Camilo Fosco*, Anelise Newman*, Pat Sukhum, Yun Bin Zhang, Nanxuan Zhao, Aude Oliva, and Zoya Bylinskii

Code, data, and models available at

Multiduration-Saliency.csail.mit.edu

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

How much time do you have? Modeling multi-duration saliency

Camilo Fosco*, Anelise Newman*, Pat Sukhum, Yun Bin Zhang, Nanxuan Zhao, Aude Oliva, and Zoya Bylinskii

MIT CSAIL IACS Adobe

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

	Static	Temporal
Individual	Fixation	Scanpath
Population	Conventional Saliency	**Multi-duration saliency**

Data collection

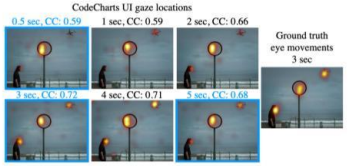
We use the **CodeCharts** interface to collect gaze fixations at precise viewing durations.

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

Try out the interface!

CodeCharts1k

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



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.

Applications

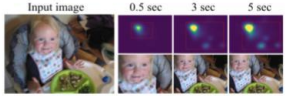
Captioning

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

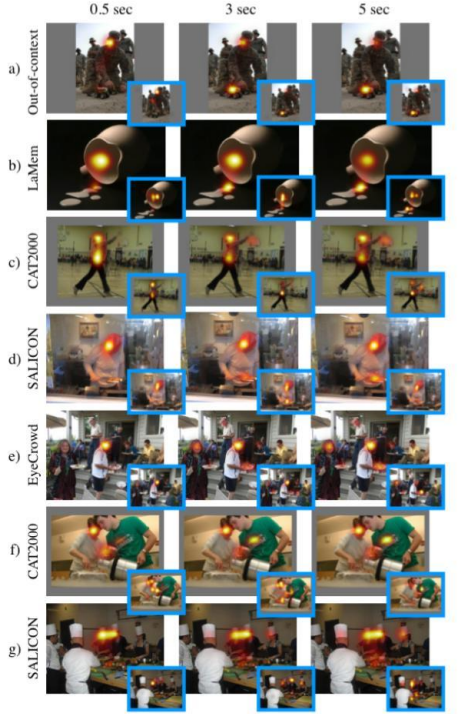
Results

CodeCharts1k

Model	NSS	CC	KL
SAMx3	2.708	0.734	0.483
SAM-MD	2.739	0.753	0.458
MD-SEM (Ours)	**2.915**	**0.765**	**0.430**

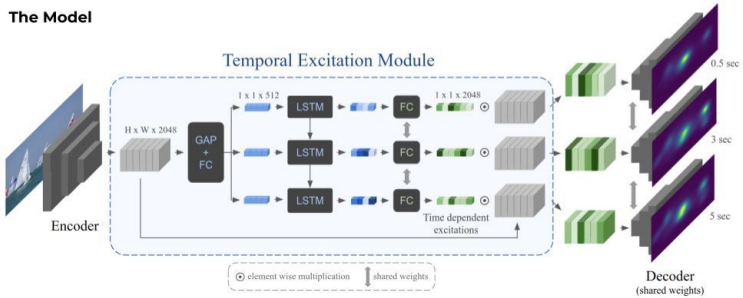
SALICON

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



Modeling multi-duration saliency

The Model

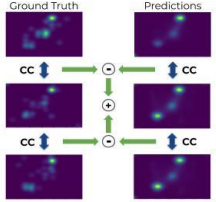


The Temporal Excitation Module

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.