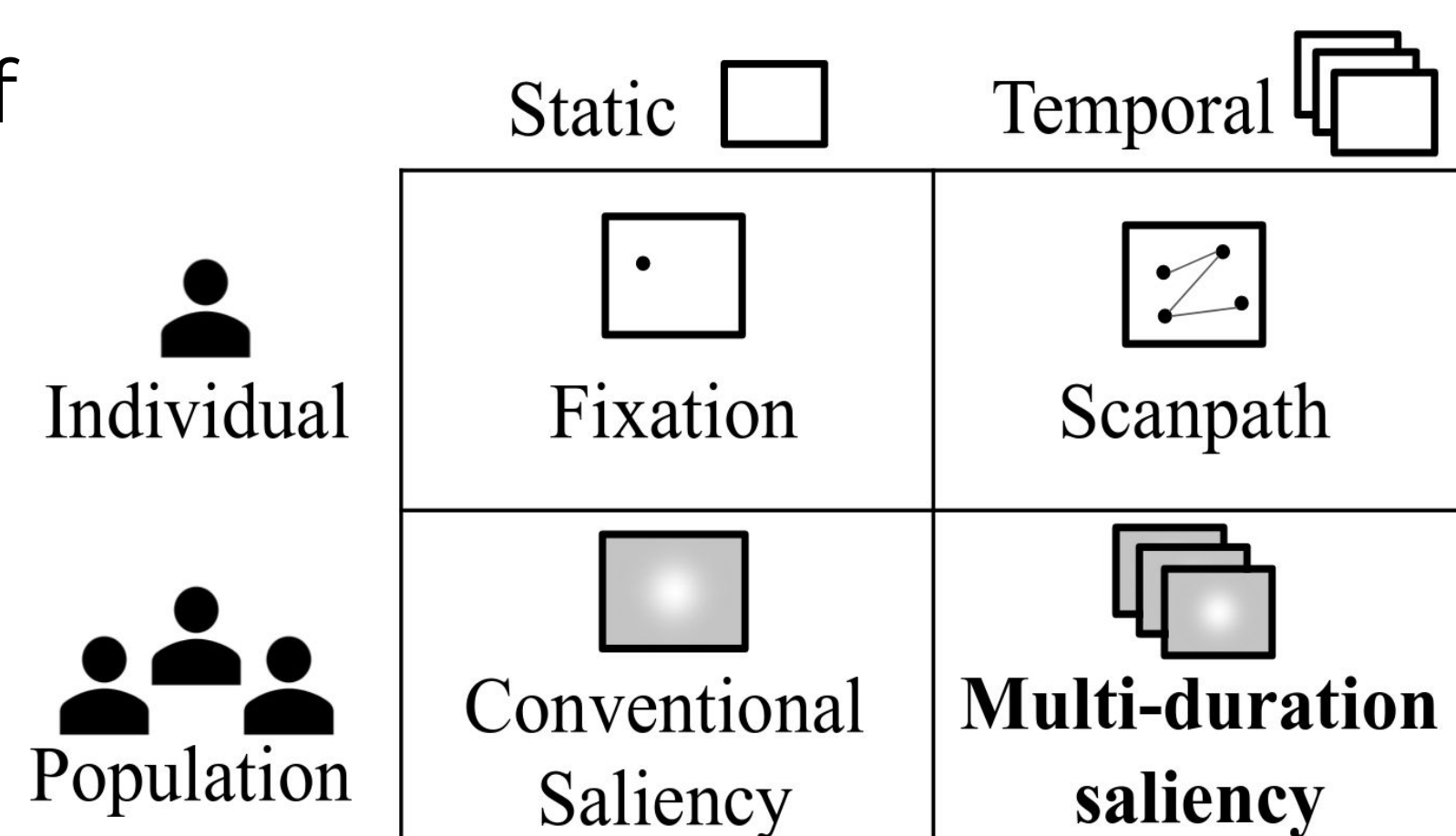


What is multi-duration saliency?

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

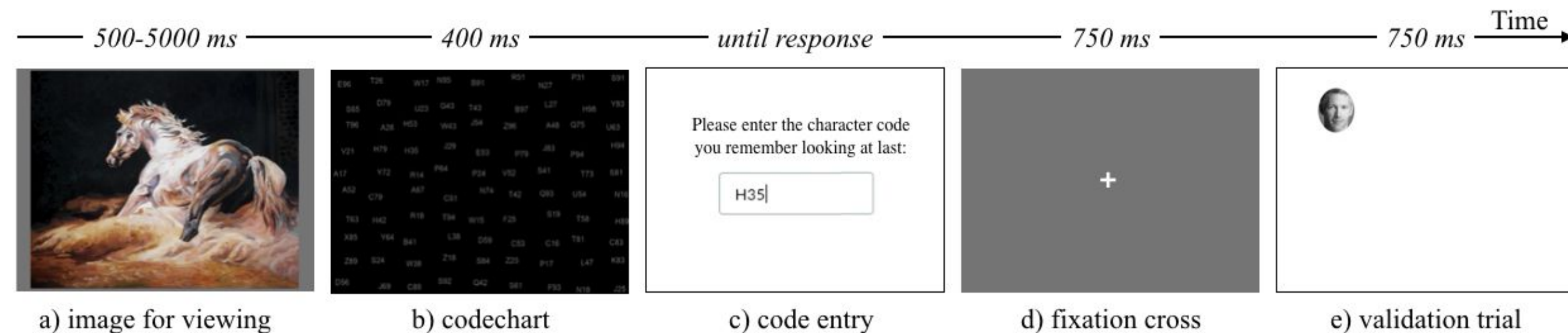
- Has the generalizability of a population-level metric
- Contains temporal information
- Easy to collect and crowdsource



Data collection

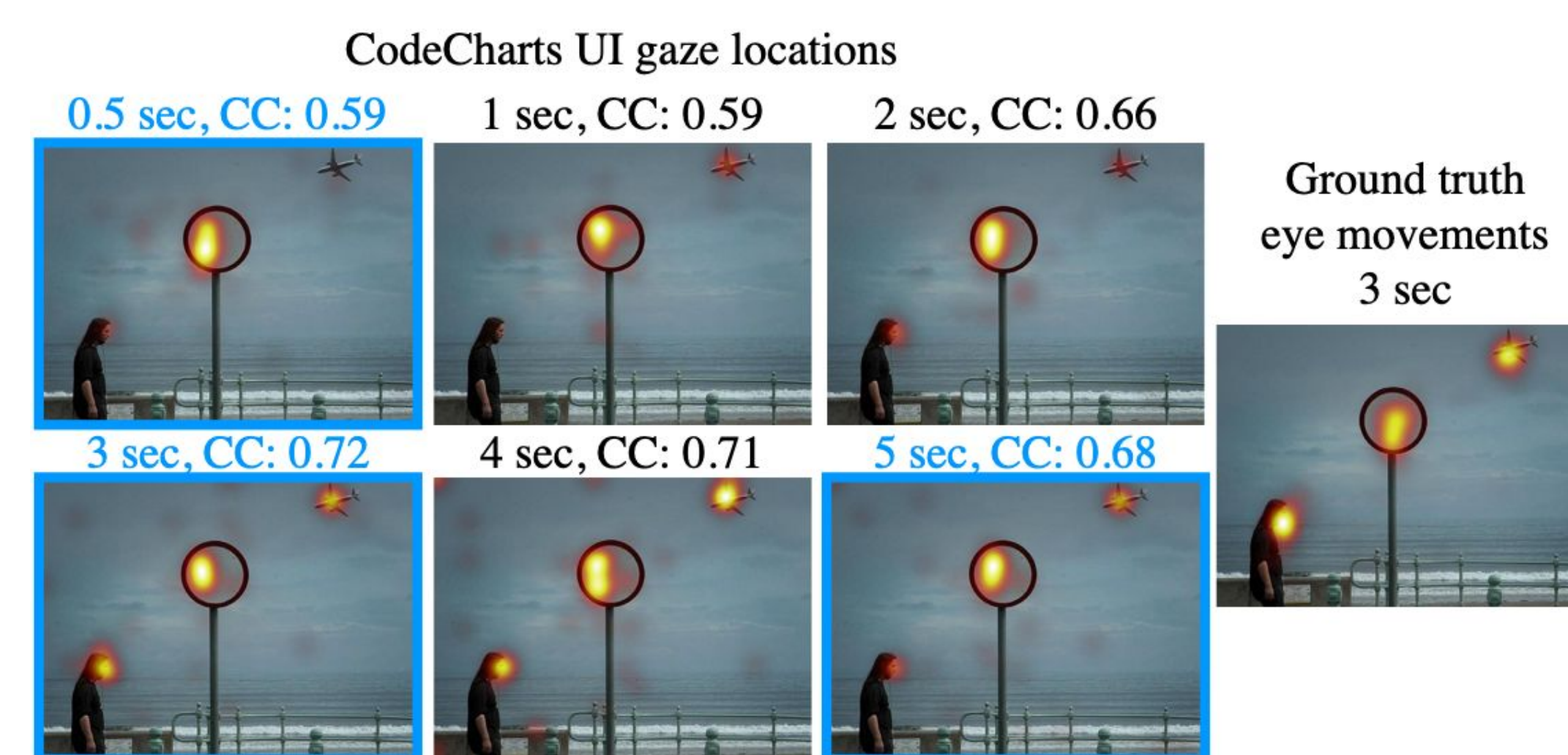
We use the **CodeCharts** interface to collect gaze fixations at precise viewing durations.

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



CodeCharts1k

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



Temporal patterns in face saliency

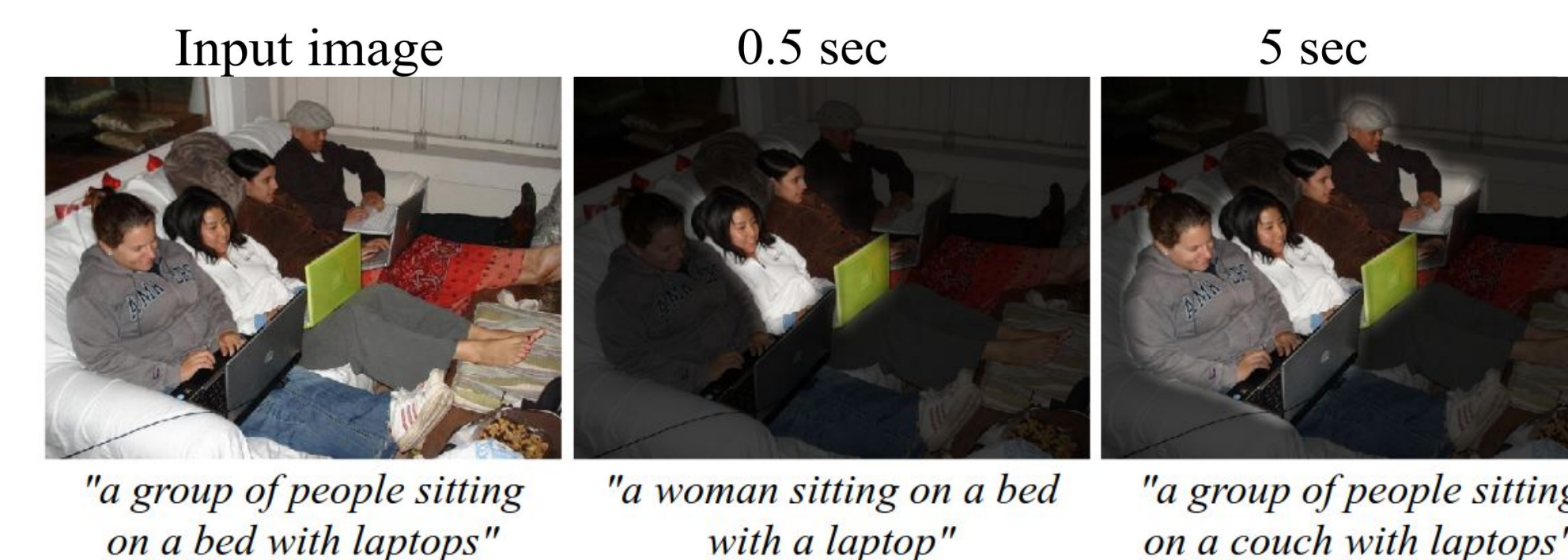


The **boomerang** pattern: attention moves away from faces at 3s and back to faces at 5s.

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

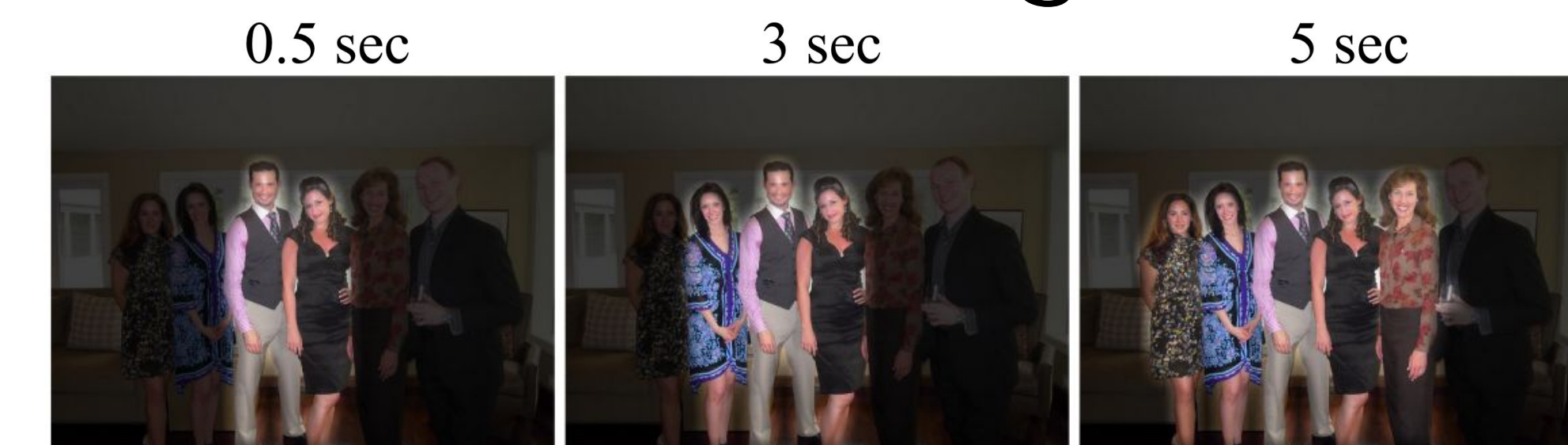
Applications

Captioning



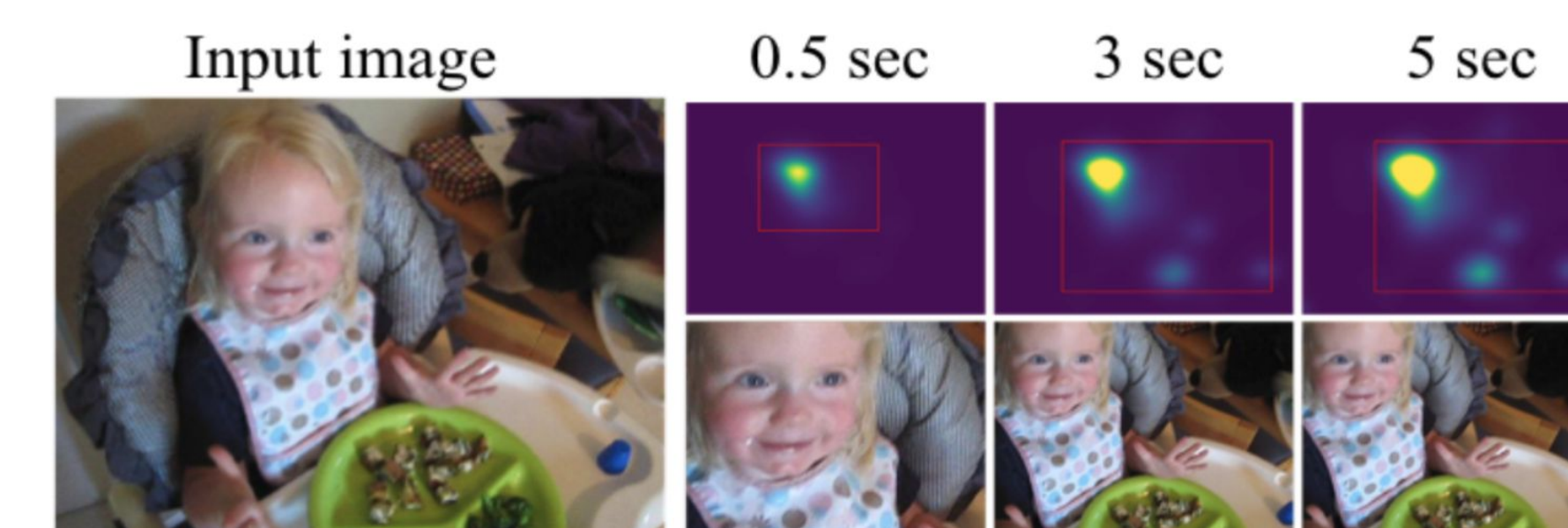
Focus a captioning module on content that is salient at different durations

Rendering



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

Cropping



Generate image thumbnails/summaries tailored to a certain duration

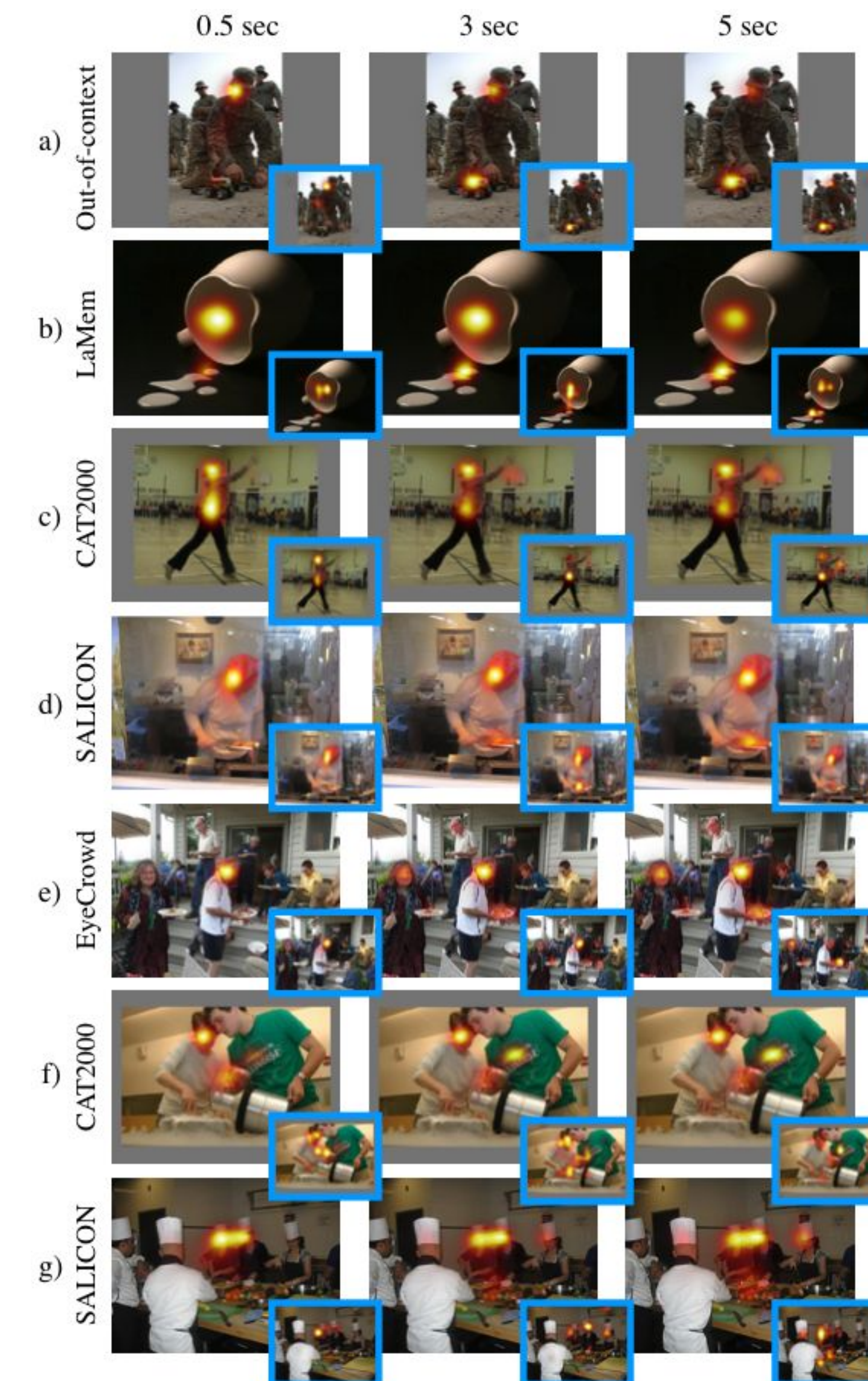
Results

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

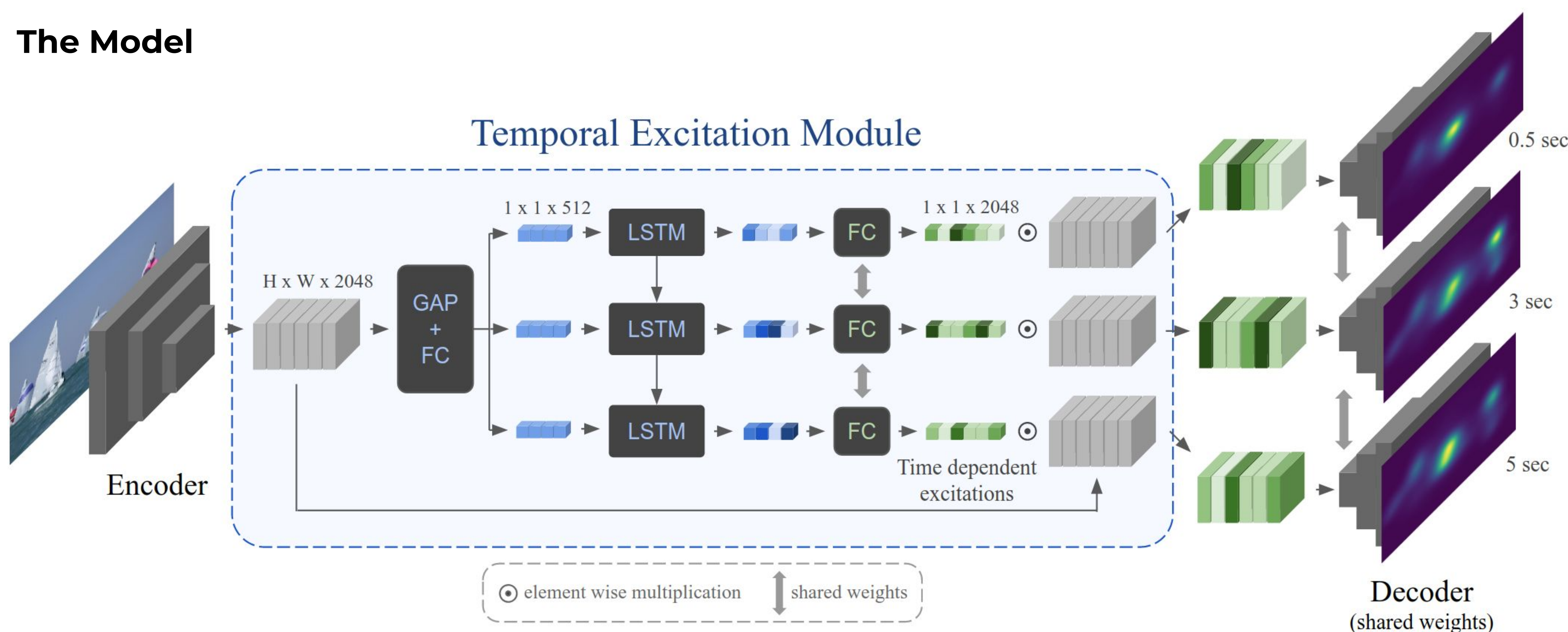
SALICON

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

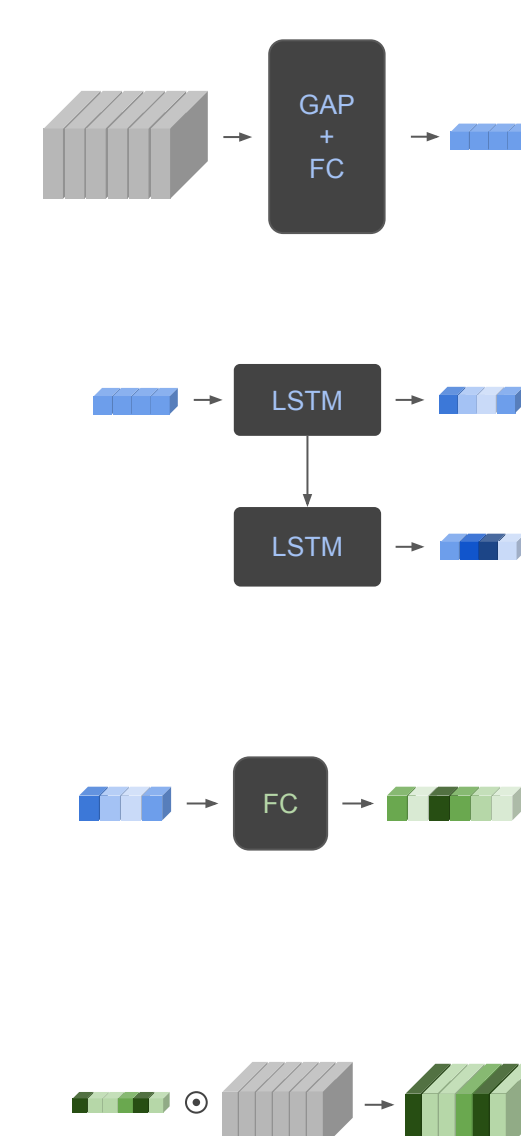
The Model



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

