TempSAL - Uncovering Temporal Information for Deep Saliency Prediction

Bahar Aydemir, Ludo Hoffstetter, Tong Zhang,

Mathieu Salzmann and Sabine Süsstrunk

https://ivrl.github.io/Tempsal

Paper tag: TUE-PM-223

Background : Saliency prediction

Input image

Saliency prediction network

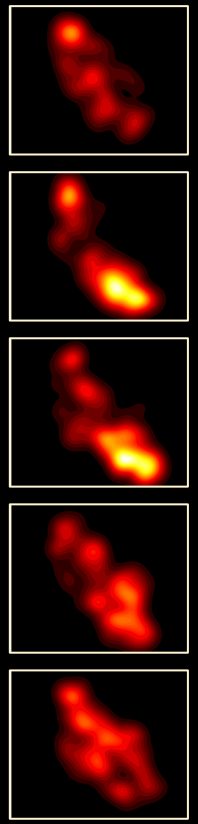
Prediction

Background : Saliency prediction

Temporal Saliency Prediction

Input image

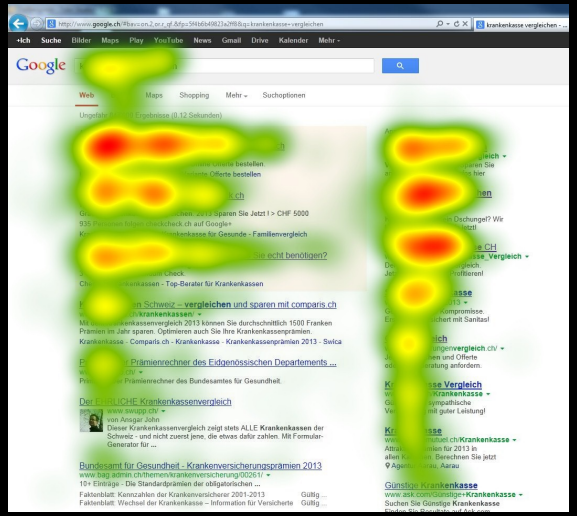
TempSAL



Temporal saliency predictions

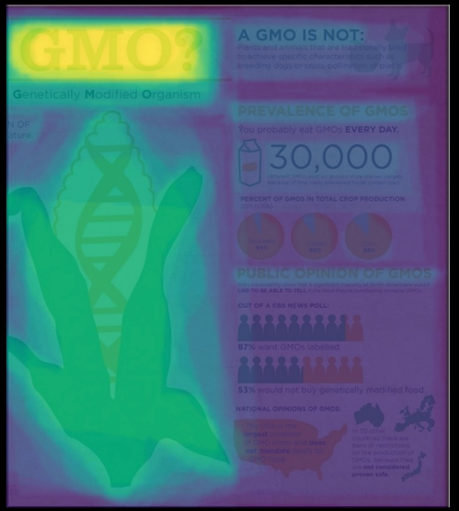
Image saliency prediction

Applications



Website design

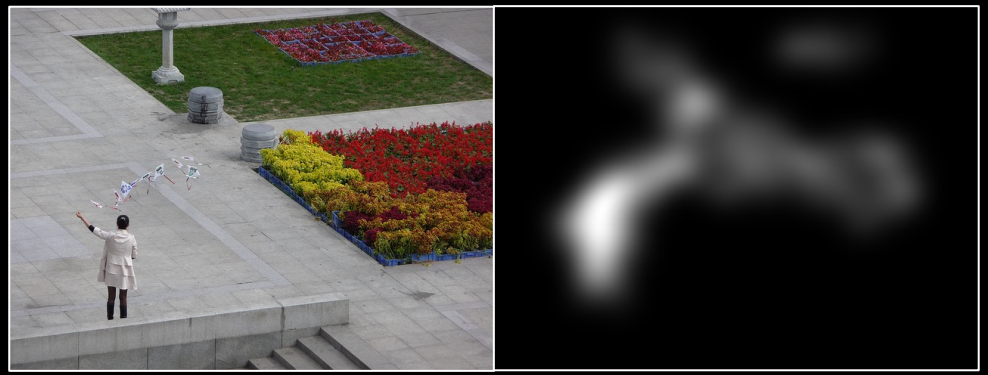
Advertising



Infographics

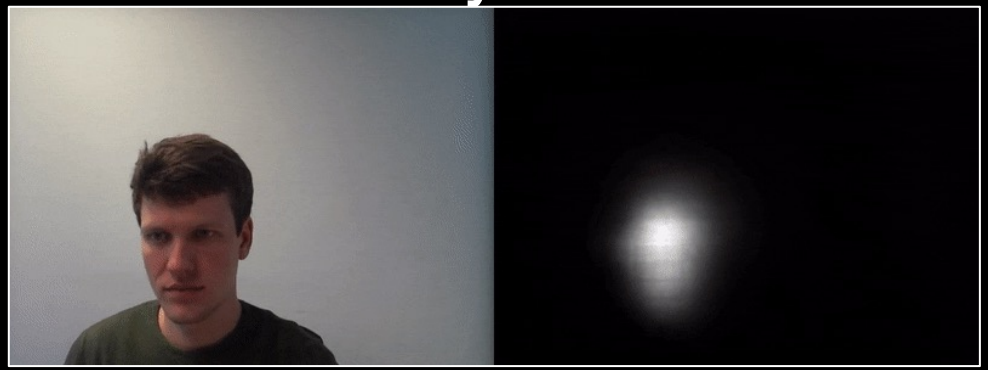
Images from: Bunnyfoot research, Thinkeyetracking, IMP1K dataset UIST 2020.

Image Saliency



A single image, a single saliency map

Video Saliency



A video (multiple frames), one saliency map per frame

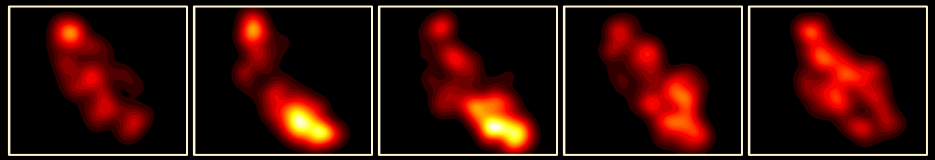
Temporal Saliency

A single image, multiple saliency maps

Kroner et al., Contextual Encoder-Decoder Network for Visual Saliency Prediction, Neural Networks, 2020

Temporal Saliency

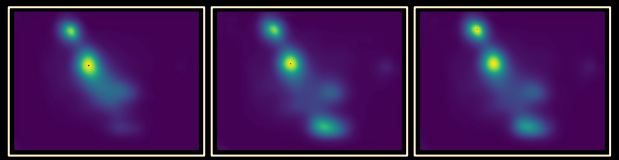
A single image, multiple saliency maps



$\mathcal{T}_1 = 0-1s$ $\mathcal{T}_2 = 1-2s$ $\mathcal{T}_3 = 2-3s$ $\mathcal{T}_4 = 3-4s$ $\mathcal{T}_5 = 4-5s$

Multi-Duration Saliency

A single image, multiple saliency maps



0 – 0.5 s 0 – 3 s 0 – 5 s

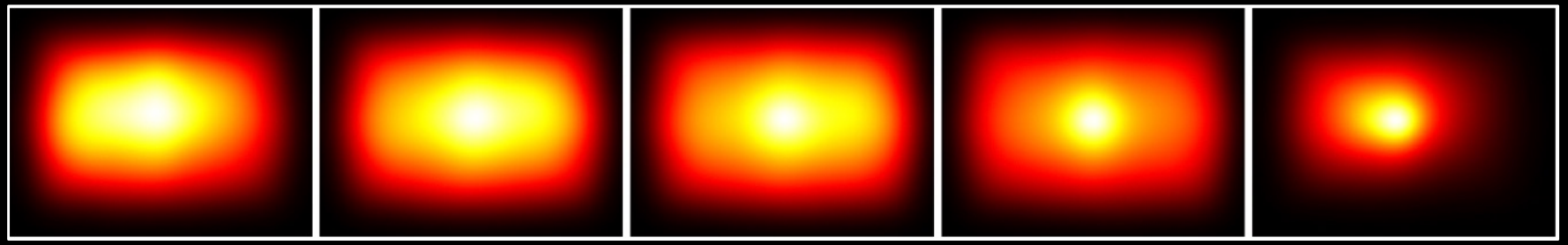
Fosco et al., How much time do you have? Modeling multi-duration saliency, CVPR 2020.

Temporal patterns

Jiang et al., SALICON: Saliency in context, CVPR, 2015.

Temporal patterns in the dataset

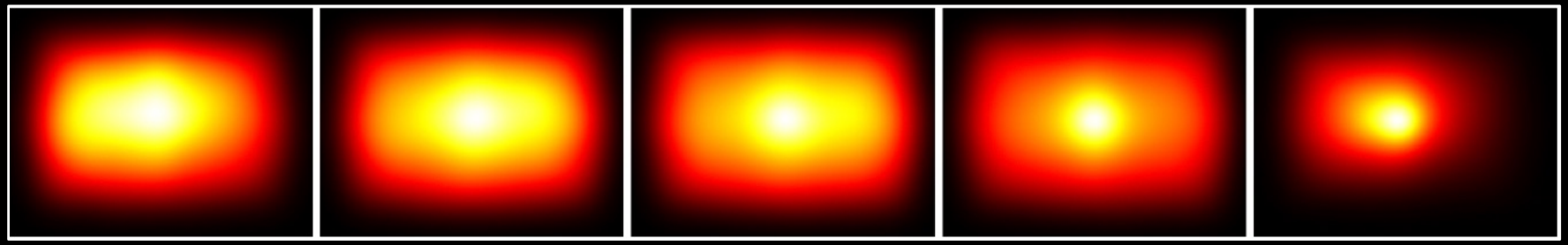
Average slices



$\mathcal{T}_1 = 0 - 1s$ $\mathcal{T}_2 = 1 - 2s$ $\mathcal{T}_3 = 2 - 3s$ $\mathcal{T}_4 = 3 - 4s$ $\mathcal{T}_5 = 4 - 5s$

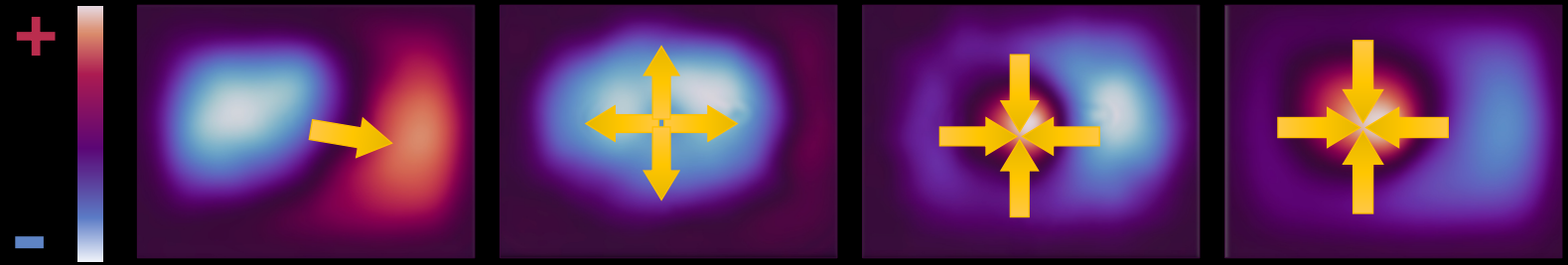
Temporal patterns in the dataset

Average slices



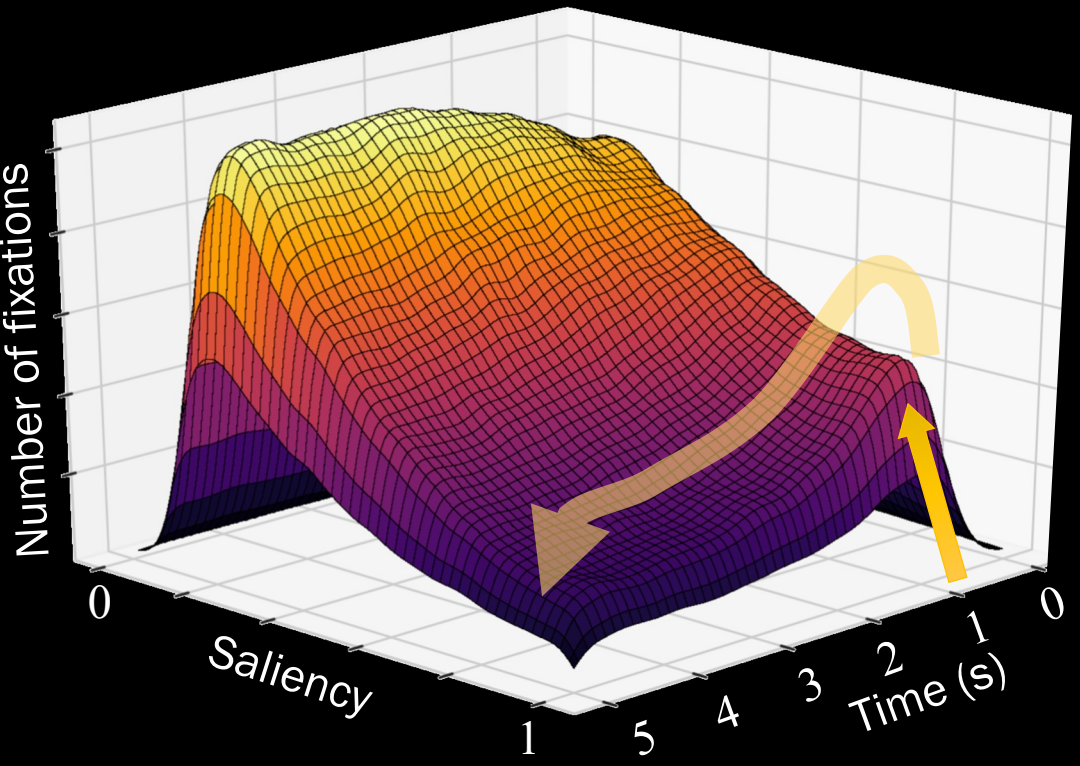
$\mathcal{T}_1 = 0 - 1s$ $\mathcal{T}_2 = 1 - 2s$ $\mathcal{T}_3 = 2 - 3s$ $\mathcal{T}_4 = 3 - 4s$ $\mathcal{T}_5 = 4 - 5s$

Difference slices



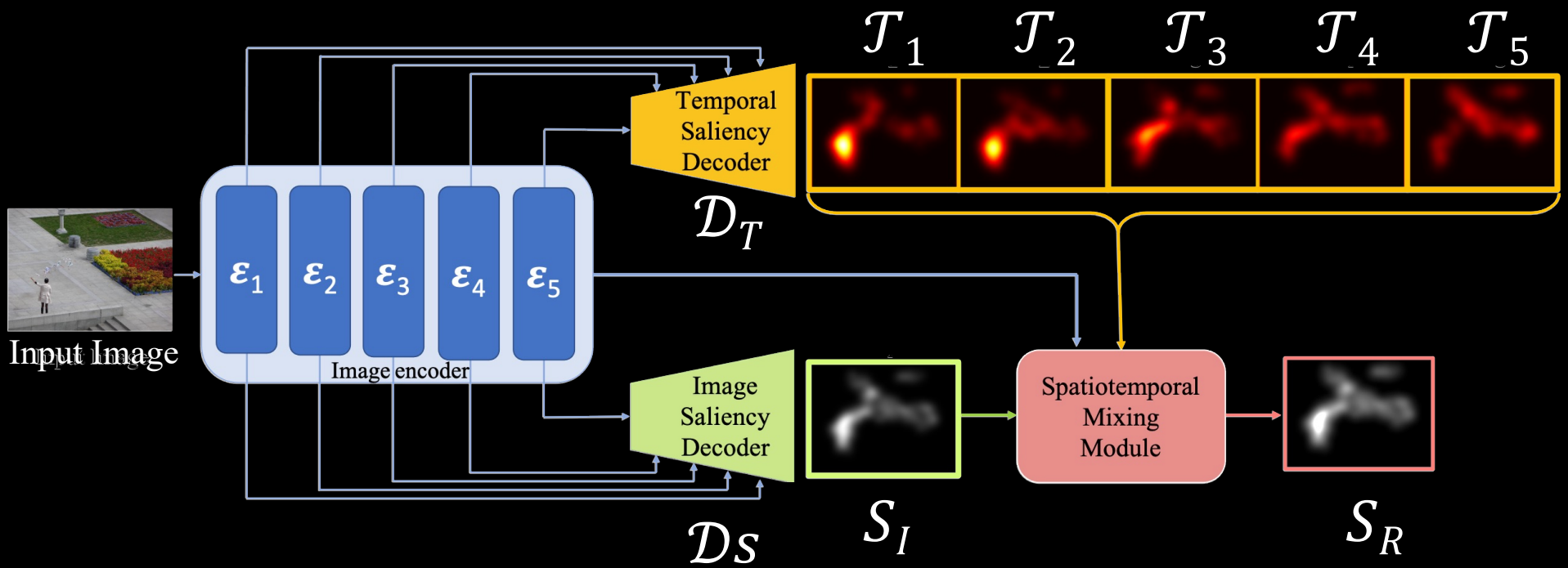
$\mathcal{T}_2 - \mathcal{T}_1$ $\mathcal{T}_3 - \mathcal{T}_2$ $\mathcal{T}_4 - \mathcal{T}_3$ $\mathcal{T}_5 - \mathcal{T}_4$

Fixations vs Time vs Saliency

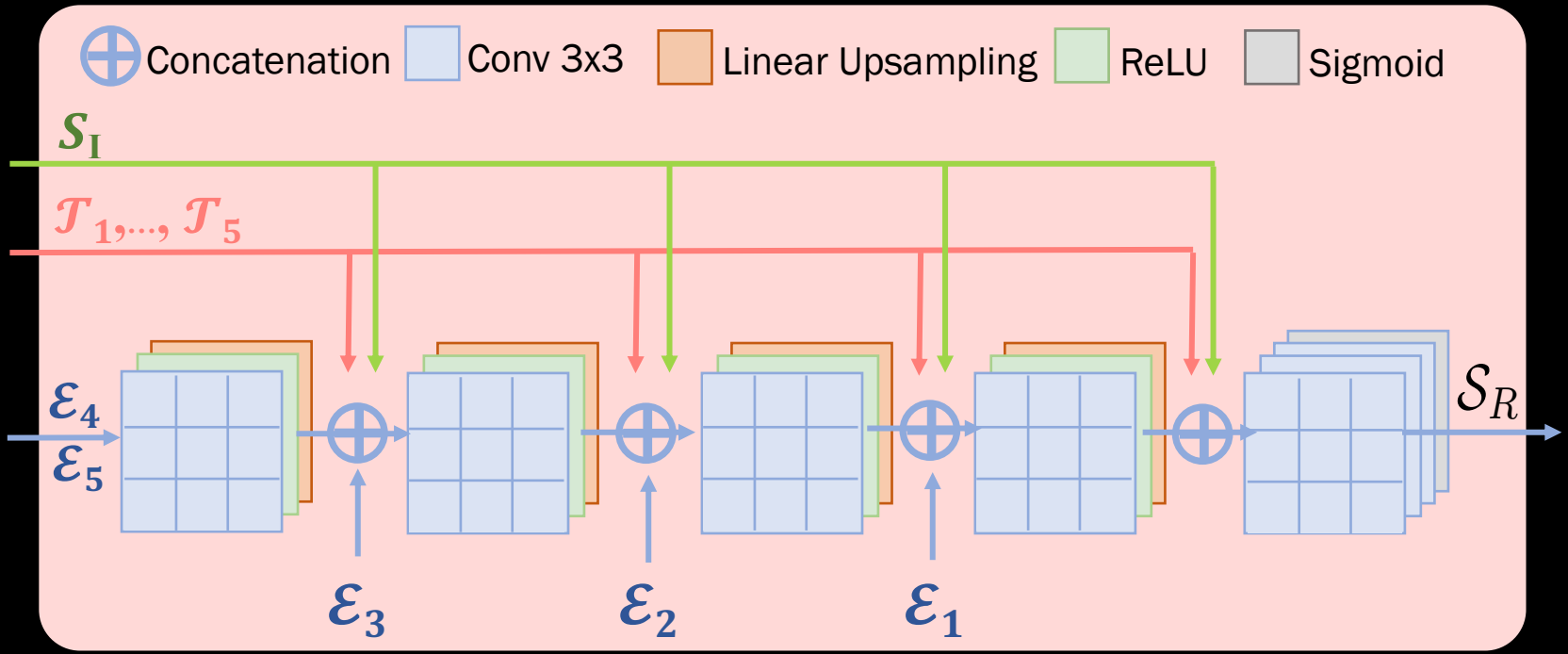


Method

Overview of our model



The Spatiotemporal Mixing Module



Qualitative results

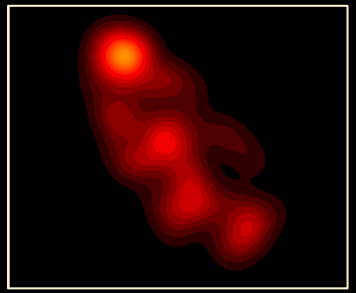
Jiang et al., SALICON: Saliency in context, CVPR, 2015.

Input image

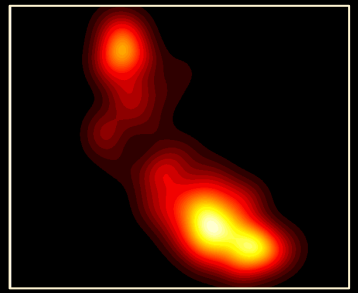
Image saliency

0 – 5 s

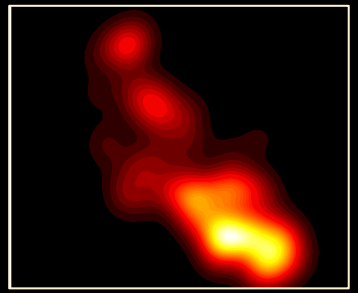
Temporal saliency



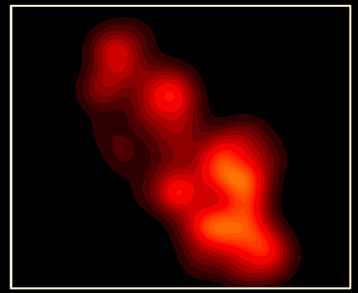
$\mathcal{T}_1 = 0 - 1s$



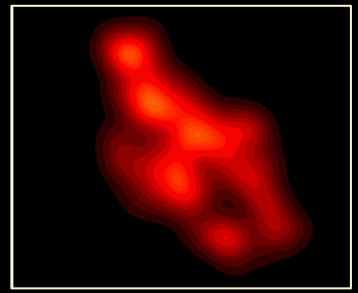
$\mathcal{T}_2 = 1 - 2s$



$\mathcal{T}_3 = 2 - 3s$



$\mathcal{T}_4 = 3 - 4s$



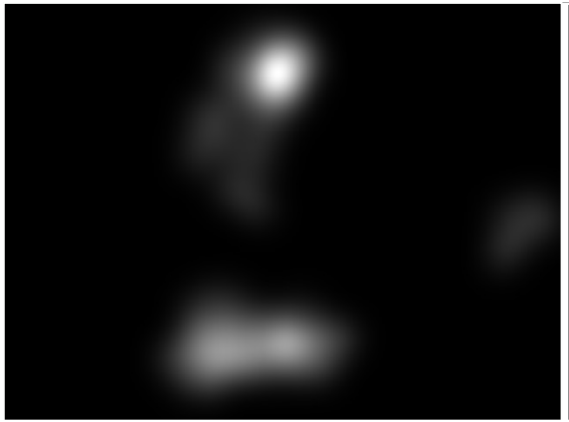
$\mathcal{T}_5 = 4 - 5s$

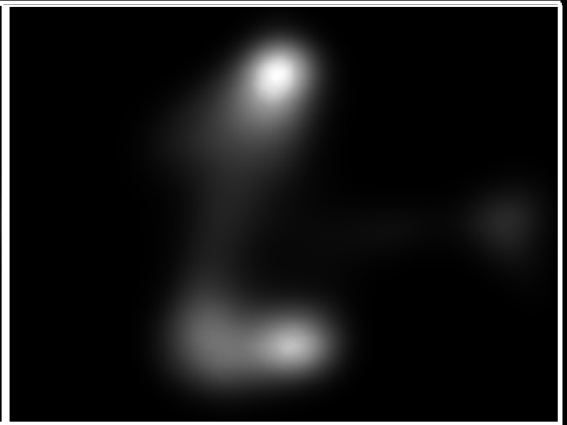
Input image

Ground truth

TempSAL

Image Saliency

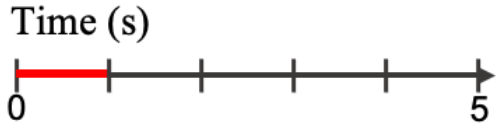


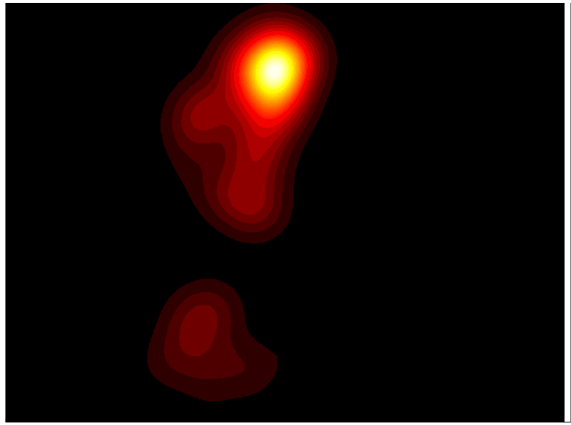


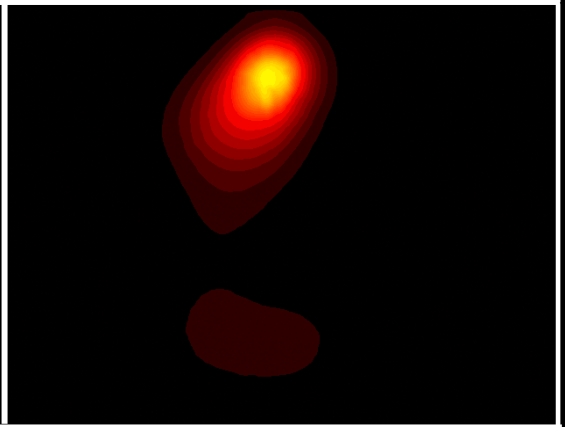
Temporal Saliency

$\mathcal{T}_1$

0-1 s





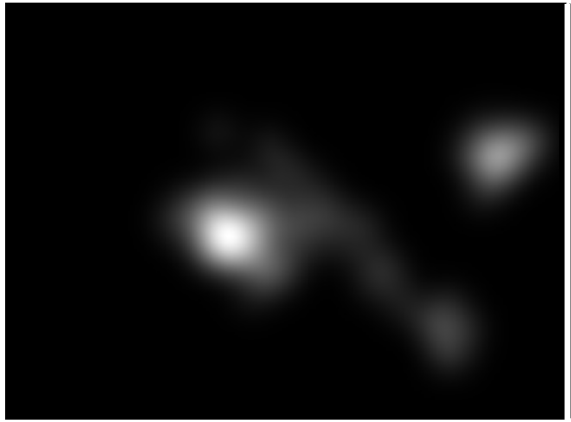


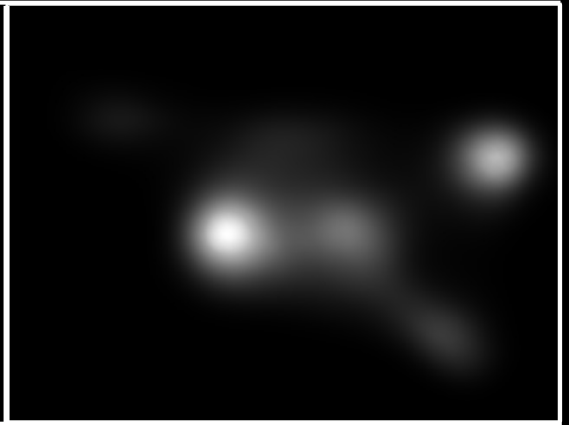
Input image

Ground truth

TempSAL

Image Saliency

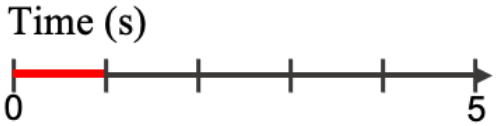


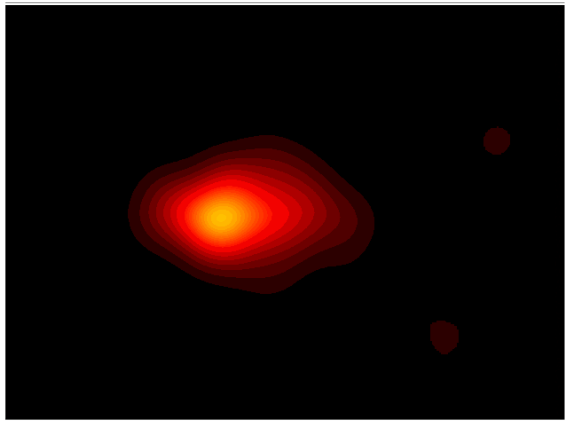


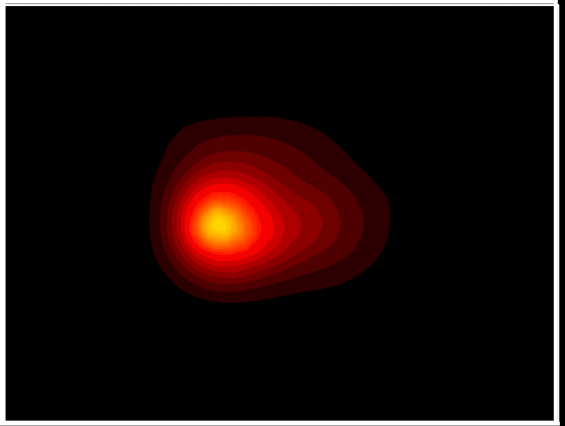
Temporal Saliency

$\mathcal{T}_1$

0-1 s





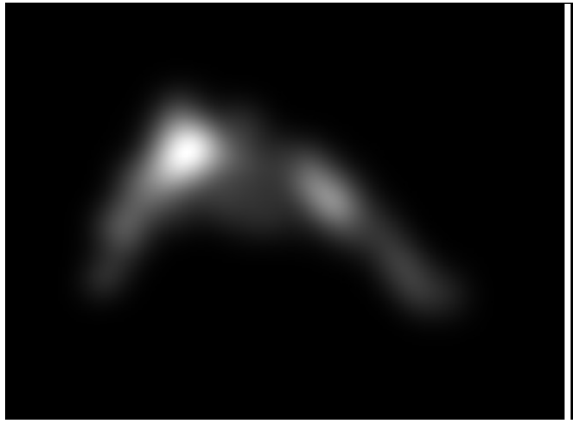


Input image

Ground truth

TempSAL

Image Saliency



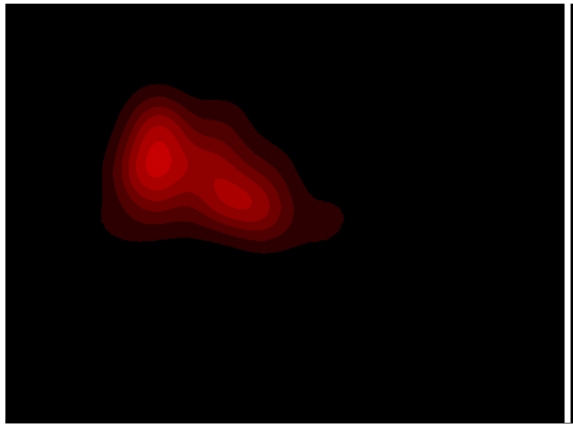
Temporal Saliency

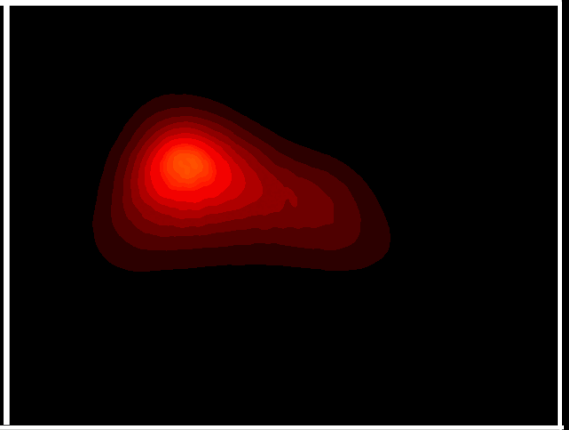
$\mathcal{T}_1$

0-1 s

Time (s)

0 5



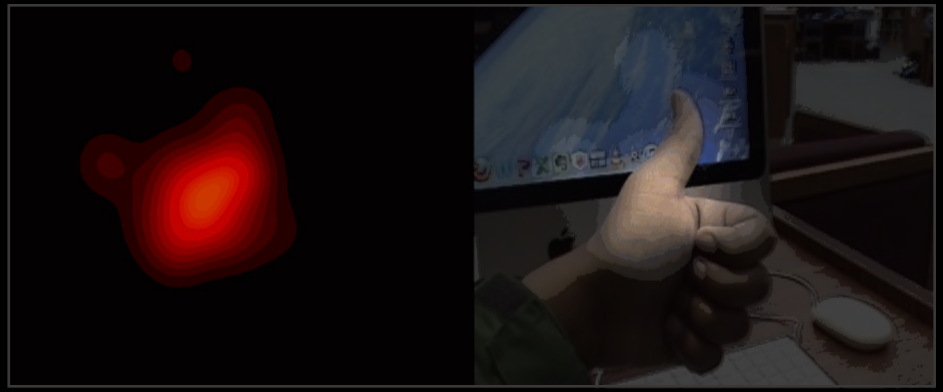


Quantitative results – Image saliency
SALICON dataset

Quantitative results – Multiduration saliency
CodeCharts1k dataset

Jiang et al., SALICON: Saliency in context, CVPR, 2015.
Fosco et al., How much time do you have? Modeling multi-duration saliency, CVPR 2020.

Conclusion



EPFL

IVRL

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

https://ivrl.github.io/Tempsal

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

This work was supported by the Swiss National Science Foundation via the **Sinergia grant CRSII5−180359**.