

A Two-Way Street

A multidisciplinary survey about visual saliency-attention, encompassing *Film Studies*, *Gestalt Psychology*, *Philosophy of Mind*, *Neuroscience*, *Neuroaesthetics*, and *Computational Vision*.

Lein de Leon Yong, Arizona State University

Background

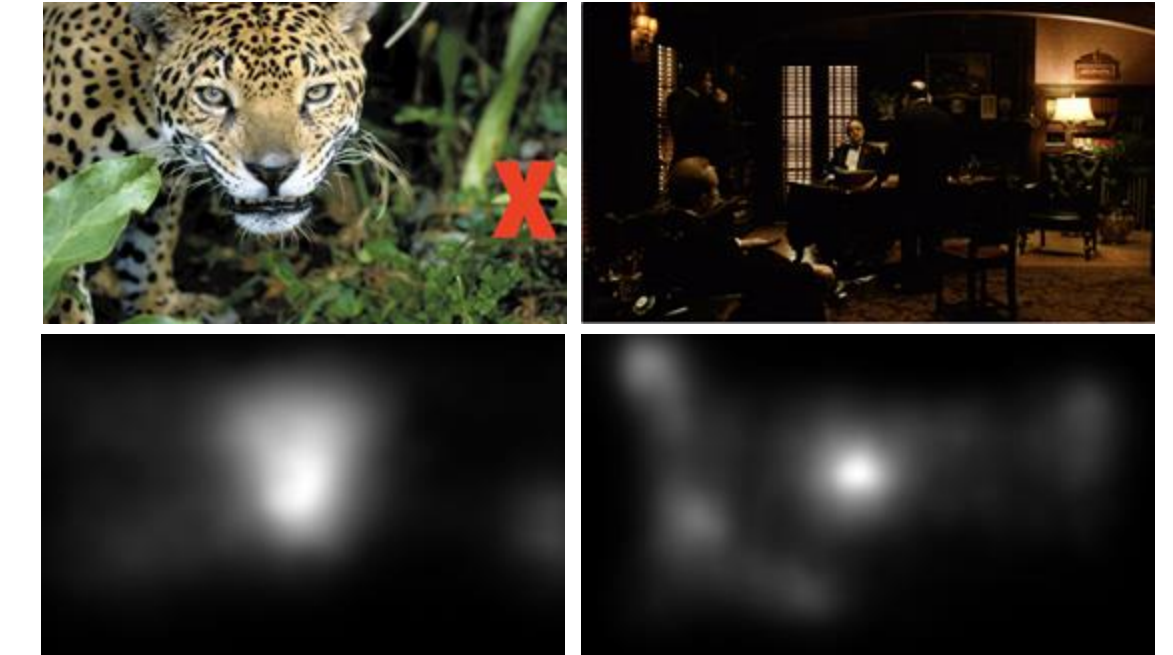
Saliency refers to the features that stand out objects through **bottom-up** processes.

Attention involves **top-down cognitive selection** influenced by higher-level semantic factors and remains less understood due to its complexity. Integrating both saliency and attention are essential for a thorough understanding of cognitive allocation.

Objective

This literature review intends to survey how different disciplines conceptualize the selection process of relevant stimuli from an environment, tracing an arc through cognitive psychology, humanities, the arts, and computational vision, sketching a multidisciplinary understanding of saliency-attention.

Saliency mapping



Attention

- Organizes the chaotic sensations of the surrounding environment into a coherent universe of experience by selecting what is significant and consequential. (Münsterberg 1916, 72-91)
- Is a process of entire pattern and configurations, instead of components, being the whole more significant the sum of its parts. (Wagemans 2012, 1172-1217)
- A condition for perception that can be possible without the voluntary attention of the subject, however this an exception. (Hamlyn 1957, 9-11)

Saliency

- The perception of an object as separated from a continuous background. (Thom 1990, 1-17)
- A cognitive cycle that encompasses the fovea's bottom-up mechanism and top-bottom brain activity that applies cultural memory, goal analysis and other cognitive processes. (Neisser 1964, 94-103) (Tatler et al. 2001, 1-21)
- Saliency mapping is the capacity of the human brain to focus on crucial scenes to preserve its resources. (Itti and Koch 2000, 10-12)
- Automated saliency mapping is based on two tasks: eye fixation prediction and salient object detection. (Yan 2022, 1-22)
- Object detection is combined with saliency mapping to predict attention by providing semantic information. (Alexe 2010, 73-80)

Specific Salient Features

- Include portraits (faces), curved lines, motion. (Chamberlain 2023, 295-314)
- One face is more salient than two faces. (Marat 2019, 197, 209)
- Pregnant forms are salient forms of biological importance that trigger physiological reactions related to reward or distress recall, involving anticipation and avoidance. (Thom 1990, 1-17)
- People will look at the same area where other people are looking at. (Park et al. 2013, 3503-3510)

Temporal saliency

- Change within time is essential in defining what is salient. A method to map saliency through measuring pixel value changes between the frames of a video. (Zhai and Shah 2006, 815, 824)

Film Studies

- When engaging with art, we move our attention from the practical outcome of perception to the actual process itself. (Bordwell 1985, 29-47)
- Salient characteristics of the artwork prompt the viewer to formulate perceptual hypotheses and construct a tale. (Bordwell 1985, 29-47)
- Films can be understood as instruments created purposely to guide viewers' attention through salient elements eliciting emotional, narrative, and descriptive meaning. (Carroll and Seely 2013, 53-75)
- Films catch our attention through temporal and spatial changes, since it is easier to perceive variations as salient than consistency. (Bordwell, Thompson 2001, 189)

References

