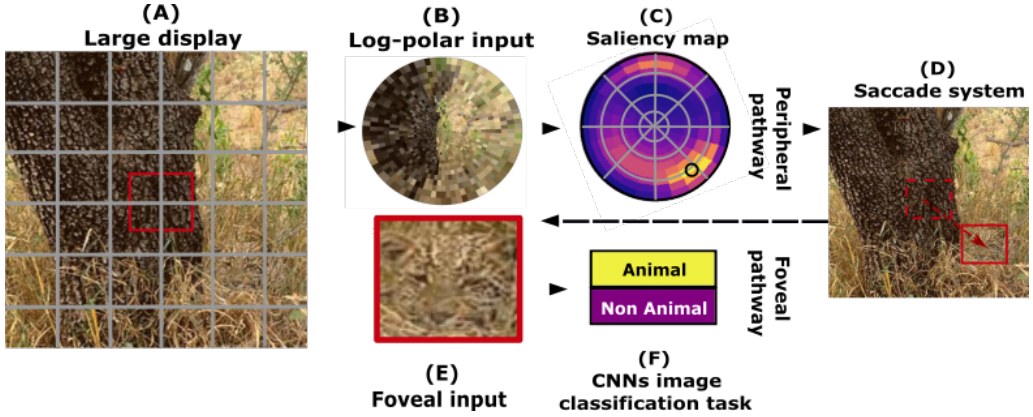
# ULTRA-RAPID VISUAL SEARCH IN NATURAL IMAGES USING ACTIVE DEEP LEARNING

*Jean-Nicolas Jeremie* [1,*], *Emmanuel Dauce* [1,2], *Laurent Perrinet* [1]

[1] Institut de Neurosciences de la Timone, CNRS/Aix-Marseille Univ, France.
[2] Ecole Centrale, Marseille, France
[*] jean-nicolas.jeremie@etu.univ-amu.fr

Visual search, that is, the simultaneous localization and detection of a visual target of interest, is a vital task. Applied to the case of natural scenes, searching for example to an animal (either a prey, a predator or a partner) constitutes a challenging problem due to large variability over numerous visual dimensions such as shape, pose, size, texture or position. Yet, biological visual systems are able to perform such detection efficiently in briefly flashed scenes and in a very short amount of time [1].Deep convolutional neuronal networks (CNNs) were shown to be well fitted to the image classification task, providing with human (or even super-human) performance. Previous models also managed to solve the visual search task, by roughly dividing the image into sub-areas. This is at the cost, however, of computer-intensive parallel processing on relatively low-resolution image samples. Taking inspiration from natural vision systems, we develop here a model that builds over the anatomical visual processing pathways observed in mammals, namely the “What” and the “Where” pathways [2]. It operates in two steps, one by selecting regions of interest, before knowing their actual visual content, through an ultra-fast/low resolution analysis of the full visual field, and the second providing a detailed categorization over the detailed “foveal” selected region attained with a saccade.




Here, the peripheral pathway (top row) is applied to a large display from a natural scene (A): It is first transformed into a retinotopic log-polar input (B) and we then learn to return a “saliency map” (C). The latter infers, for different positions in the target, the predicted accuracy value that can be reached by the foveal pathway, mimicking the “Where” pathway used for global localization. The position with the best accuracy will feed a saccade system (D), adjusting the fixation point at the input of the foveal pathway (bottom row). It takes a subsample (E) of the large display (A), over which a categorization is done (F), mimicking the “What” pathway. Modeling this dual-pathways architecture allows to offer an efficient model of visual search as active vision. In particular it allows to fill the gap with the shortcomings of CNNs with respect to physiological performances. In the future, we expect to apply this model to better understand visual pathologies in which there would exist a deficiency of one of the two pathways.

## References

1. Thorpe, Fize, Marlot, 1996, *Nature* 381(6582):520-522, 10.1038/381520a0

2. Dauce, Perrinet, 2020, *J Vision* 1326:165-178, 10.1007/978-3-030-64919-7_17