

Visual search as active inference

No Author Given

No Institute Given

Abstract. Visual search is an essential cognitive ability, offering a prototypical control problem to be addressed with Active Inference. Under a Naive Bayes assumption, the maximization of the information gain objective is consistent with the separation of the visual sensory flow in two independent pathways, namely the “What” and the “Where” pathways. On the “What” side, the processing of the central part of the visual field (the fovea) provides the current interpretation of the scene, here the category of the target. On the “Where” side, the processing of the full visual field (at lower resolution) is expected to provide hints about future central foveal processing. A map of the classification accuracies, as obtained by counterfactual saccades, defines a utility function on the motor space, whose maximal argument prescribes the next saccade. The comparison of the foveal and the peripheral predictions finally forms an estimate of the future information gain, providing a simple and resource-efficient way to implement information gain seeking policies in active vision. This dual-pathway information processing framework is found efficient on a synthetic visual search task and we show here quantitatively the role of the precision encoded within the accuracy map. More importantly, it is expected to draw connections toward a more general actor-critic principle in action selection, with the accuracy of the central processing taking the role of a value (or intrinsic reward) of the previous saccade.

Keywords: Object detection · Active Inference · Visual search · Visuomotor control · Deep Learning.

1 Problem statement: formalizing visual search as accuracy seeking

Moving fast the eye toward relevant regions of the scene, in order to minimize the surprise, interestingly combines elements of action selection (moving the eye) with visual information processing, making it an ideal test-bed for implementing the Active Inference (AI) principles and in particular for the active sampling hypothesis [PAF14]. During natural selection, the visual sensors have evolved toward maximizing their efficiency under strong energy constraints. Vision in most mammals, for instance, has evolved toward a foveated sensor, maintaining a high density of photoreceptors at the center of the visual field, and a much lower density at the periphery. This limited bandwidth transmission is combined with a high mobility of the eye, that allows to displace the center of sight toward different parts of the visual scene, at up to 900 deg/s in humans. This action perception loop uniquely specifies an AI problem [Fri+12; Dau18].

In particular, we will focus here on *visual search* [Eck11] which is the cognitive ability to locate a visual object in cluttered visual scene by placing the fovea on the object, in order to identify it. As such, visual search intimately links the sampling of visual space (as it is done by the sensory apparatus) to the behavior which directs this sampling through the action of moving the direction of gaze. Note that the retina is mostly sampling visual information on the fovea, and it is likely that the target lies in the periphery where acuity is less precise. It is therefore probable that the target is not identifiable with the current information contained on the retinal image. As a consequence, visual search involves the problem that, given a limited observability, the object has to be localized *before* being identified.

This observation highlights an important hypothesis for solving the visual search problem. Indeed, it is obvious that the position of an object is *not* independent of its identity. The semantic content of a visual scene is indeed defined by the positions and identities of the many objects that it contains, and those objects are not exchangeable. We however consider here a counter-intuitive assumption, that is that the visual system of mammals is built around such an independence hypothesis. This hypothesis tells, in short, that the identity and the position can be processed independently and, more importantly, should be processed *sequentially* to arrive at a correct identification. This assumption, largely exploited in machine learning, is known as the “Naïve Bayes” assumption. It simply considers as independent the different factors (or latent features) that explain the data. Selecting an object and identifying both its position and category may thus be the elementary brick of visual processing. It may moreover explain the general separation of visual processing into a ventral and a dorsal pathways. These specific processing pathways are then devoted to the processing of the visual flow, either to identify the semantic content of the visual field (the “What” pathway), or to decide where to orient next the line of sight (the “Where” pathway), in a continual and incremental turn-taking process, contributing to understand and exploit the visual environment in the most efficient way.

In order to implement those principles into a concrete image processing task, we construct a simple yet ecological virtual experiment: After a fixation period of 200 ms, an observer is presented with a luminous 128×128 display showing a single target overlaid on a realistic noisy background (see Figure 1-A). This target is drawn in our case from the MNIST database of digit images of size 28×28 [Lec+98]. This image is displayed for a short period of about 500 ms which allows to perform at most one saccade toward the target. The goal of the agent is to correctly identify the digit. Let us define as $x = (x_1, x_2)$ the possible

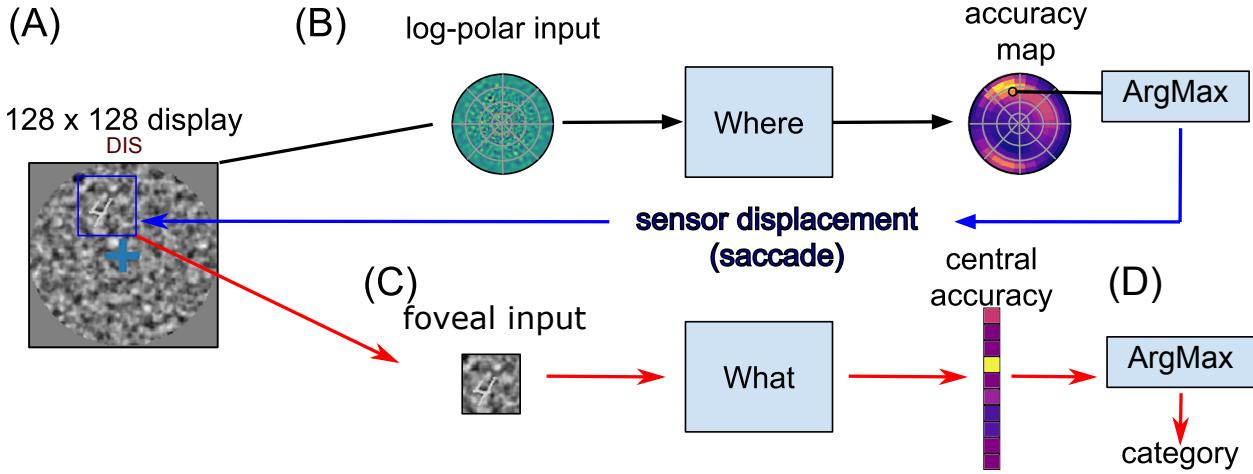


Fig. 1. Computational graph. Based on the anatomy of mammals' visual pathways, we define the following stream of information to implement visual search, one stream for localizing the object in visual space ("Where?"), the other for identifying it ("What?"). **(A)** The visual display is a stack of three layers: first a natural-like background noise is generated, characterized by noise contrast, mean spatial frequency and bandwidth [San+12]. Then, a sample digit is selected from the MNIST dataset [Lec+98], rectified, multiplied by a contrast factor and overlaid at a random position. Last, a circular, gray mask is put on. **(B)** The visual display is then transformed in a retinal input which is fed to the "Where" pathway. This observation is generated by a bank of filters whose centers are positioned on a log-polar grid and whose size increases proportionally with the eccentricity. The "Where" network outputs a collicular-like accuracy map. It is implemented by a three-layered neural network consisting of the retinal log-polar input, two hidden layers (fully-connected linear layers combined with a ReLU non-linearity) with 1000 units each. This map has a similar log-polar (retinotopic) organization and predicts the accuracy at the counter-factual positions of affordable saccades. The position of maximal activity in the "Where" pathway serves to generate a saccade s which displaces the center of gaze at a new position $x = \rho(s)$. **(C)** This generates a new sensory input $f(x)$ in the fovea which is fed to a classification network ("What" pathway). This network is implemented using the three-layered LeNet neural network [Lec+98]. This network outputs a vector predicting the accuracy of detecting the correct digit. **(D)** Depending on the (binary) success of this categorical identification, we can reinforce the What networks, by supervisedly learning to associate the output with the ground truth through back-propagation, and then the "Where" network by updating the approximation of the accuracy map.

bi-dimensional retinocentric positions of the object (that is, $x \stackrel{\text{def}}{=} x_0 = (0,0)$ at the center of gaze) and by $k \in \{0, \dots, 9\}$ its potential identities. At any given trial t drawn from the set \mathcal{T} of trials of our virtual experiment, the problem of identifying the object is solved for instance by extracting the 28×28 image $f^t(x)$ around a potential position x and by using a deep neural network [Lec+98] to infer its category. This network returns a multinomial distribution vector $a_k(f^t(x)) \in \mathbb{R}^{10}$ (with $\sum_k a_k(f^t(x)) = 1$). This network takes here the role of the "What" pathway.

Knowing $f^t(x)$, the correctness of the response $T(f^t(x))$ as we test our model (a value 1 for a correct detection $k^t = \arg \max_k a_k(f^t(x))$ or else 0) can be regarded as a binary random variable. This binary variable can be sampled at different x , with different success or failures depending on the position. It is essential for visual search to estimate the correctness of the test *prior* to a saccade, that is, to predict the statistics of $T(f^t(x))$. Usefully, the "What" neural network outputs the estimate of the chance of success for every possible responses, such that the *accuracy* at x is simply:

$$A^t(x) \approx \max_k a_k(f^t(x))$$

In principle, one could extract all possible sub-images $f^t(x)$ at all positions x , and estimate $A^t(x)$ directly. Moving the eye toward $x^t = \arg \max_x A^t(x)$ and finding the object's identity at location x would solve the problem of both identifying and locating the target. This brute-force solution is of course computationally prohibitive, but provides a baseline toward a more biologically-relevant processing. Instead, the eye next position x is a motor displacement that should be governed by a *policy*, i.e. a method that selects the next movement from the available visual input. The set of all possible displacements forms a motor map, and such a policy can be formalized as a mapping from the visual input space toward this motor map. Following the classical reinforcement learning literature, the motor map is expected to provide a value over the space of actions. We postulate here that *the value of the motor displacement toward x is identified with the classification accuracy obtained at position x* . Moreover, we will show that, with minimal simplifying assumptions, this postulate can be framed into a more general active inference framework.

2 Principles: supervised learning of action selection

2.1 Peripheral visual processing

In our model, the fovea, which constitutes the center of the retina, effectively represents $f^t(x)$, and the peripheral region provide a visual information with a decreasing precision with respect to the eccentricity. The exponential decrease of the density of photo-receptors with respect to eccentricity [Wat14] is reflected in a non-uniform sampling of the visual data. It is here implemented as a log-polar conformal mapping, as it provides a good fit with observations in mammals [JB10] and has a long history in computer vision and robotics. These coordinates are denoted as the tuple (ϵ, θ) corresponding respectively to the log-eccentricity and azimuth in (spherical) polar coordinate by $x = \rho(\epsilon, \theta) \stackrel{\text{def.}}{=} (R \cdot \exp(\epsilon) \cdot \cos \theta, R \cdot \exp(\epsilon) \cdot \sin \theta)$ with R the maximal eccentricity. Both the visual features and the output accuracy map are to be expressed in these coordinates.

On the visual side, local visual features are extracted as oriented edges as a combination of the retinotopic transform with filters resembling that found in the primary visual cortex [Fis+07]. The centers of these first and second order orientation filters are radially organized around the center of fixation, with small receptive fields at the center and more large and scarce receptive fields at the periphery. The size of the filters increases proportionally with the eccentricity. Assuming this log-polar arrangement, the resulting retinal visual data at this trial is noted as a vector o^t . On the motor side, the target accuracy map is also organized radially in a log-polar fashion, making the target position estimate more precise at the center and coarser at the periphery. This modeling choice is reminiscent of the approximate log-polar organization of the superior colliculus (SC) motor map [SN87]. Then, the possible saccade location denoted as $s \stackrel{\text{def.}}{=} (\epsilon, \theta)$. Each coordinate of the visual field, except for the center, is mapped on this saccadic motor map.

Then, the objective of the “Where” processing pathway is to find the utility function that approximates at best the accuracy map:

$$q(s|o^t) \approx A^t(\rho(s))$$

where $\rho(s)$ is the future position of gaze. As a consequence, the visual search problem can be summarized as finding the function q which would predict the accuracy of the what pathway, given the action (saccade). This prospective coding scheme thus involves forming prior beliefs about fictive, counter-factual sampling, and is thus relevant for an Active Inference treatment.

2.2 Active inference and the accuracy drive

Following the general principles expressed by Friston [Fri10], engaging in a saccade stems on maintaining the visual field within the least surprising possible state. This implies, for instance, the capability to predict the next visual input through a generative model, and orient the sight toward regions that minimize the agent’s predicted model surprise [Fri+12]. Due to their limited memory and processing capabilities, living brains do not afford to predict or simulate their sensory environment exhaustively. Given the vast diversity of possible visual fields, one should assume that only the foveated part should deserve predictive coding. This implies that the saccadic motor control should be tightly optimized in order to provide a foveal data that should allow to accurately identify (and predict) the target. Assuming the motor control is independent from the identity pathway, we take the classification success, as measured at the output of the “What” pathway, as the principal outcome of the “Where” pathway. It is assumed, in short, that the surprise should be higher in case of failure than in case of success, and that minimizing the surprise through active inference should be consistent with maximizing the likelihood of success.

For a saccade s , let o^t be the peripheral observation and $T(f^t(\rho(s)))$ the corresponding classification success.

$$\log p(T|o^t) \geq \mathbb{E}_q \log p(T|s, o^t) - \log q(s|T, o^t) + \log p(s|o^t)$$

where, at each trial t , the classification success $T(f^t(\rho(s)))$ is a realization of $p(T|s, o^t)$. In a predictive setting, this likelihood of success is precisely implemented by our accuracy map. Now, the optimization being done on s , our saccade selection process relies on maximizing the likelihood of success, i.e. $\arg \max_s p(T = 1|s, o^t)$, that is consistent with assuming that a prior is put on observing a success, whatever the saccade. From Bayes rule, we also know that optimizing on the likelihood assumes a uniform prior on action selection. Incidentally, an unimodal shape of the accuracy map indicates that a highest chance of success is found when the target is centered on the fovea, and for that reason the active inference mechanism should privilege saccades that will place the visual target at the center of the fovea. This is equivalent to identifying the location of the target in the retinotopic space, and thus inferring the spatial information from the visual field, with the future saccade taking the role of a latent variable explaining the current visual field o^t .

2.3 Learning the accuracy map

Our likelihood function p can be seen as a mapping from o^t to the set \mathcal{S} of possible saccade locations. Neural Networks are known to be in theory universal function approximators. The parametric neural network consists of a primary visual input layer, followed by two fully connected hidden layers of size 1000 with

rectified linear activation, and a final output layer with a sigmoid nonlinearity to ensure that the output is compatible with a likelihood function. In accordance with observations [CV84; SN87], a similar log-polar compression pattern is observed at the retinal input and at the motor output.

We use the BCE cost as the Kullback-Leibler divergence between the likelihood and its approximation:

$$\mathcal{L}_S^t = -T(f^t(\rho(s))) \cdot \log p(T = 1|s, o) - (1 - T(f^t(\rho(s)))) \cdot \log(1 - p(T = 1|s, o))$$

We now optimize the parameters of the neural network implementing the “Where” pathway such as to optimize the approximation made by p . This can be achieved in our feed-forward model using back-propagation with the input-output pairs (o^t, s^t) and the classification result as it is given by the “What” pathway. The role of the “What” pathway is here that of a critic of the output of the “Where” pathway (which takes the role of the actor). This separation of visuo-spatial processing into an actor and a critic is reminiscent of a more general actor-critic organization of motor learning in the brain, as postulated by [JNR02].

The natural way to collect such supervision data is to draw data one by one in our virtual experiment, iteratively generating a saccade and computing the success of the detection. This is what would be performed by an agent which would sequentially learn by trial-and-error, leading to a reinforcement scheme. To accelerate the learning in our scheme, there exists however a computational shortcut to obtain more supervision pairs. Indeed, for each input image, we know the true position x^t and identify k^t of the target. As such, one can compute the average accuracy map over the dataset and optimize equivalently

$$\mathcal{L}_S^t = - \sum_{s \in S} [A_0(x - \rho(s)) \cdot \log p(T = 1|s, o) + (1 - A_0(x - \rho(s))) \cdot \log(1 - p(T = 1|s, o))]$$

where $A_0(x - f(\rho(s)))$ is the accuracy map centered on the true position of the visual object. This accelerates learning as it scales up both the set of tested saccade positions and gives the analog bias value instead of the binary outcome of the detection.

2.4 Saccade selection

The output of the “Where” network allows to implement a function which approximates the likelihood of classification success given counterfactual saccades and the observed visual input. Then, the choice of a saccade given the likelihood may be obtained from the maximum a posteriori rule under a uniform prior on action selection

$$s^t \stackrel{\text{def.}}{=} \pi_{\max}(o^t) = \arg \max_s [q(s|o^t, T = 1)] = \arg \max_s p(T = 1|s, o^t) \cdot \Pr(s|o^t)$$

with for instance $\Pr(s|o) = \text{Unif}(s)$ an uniform prior probability on saccade selection. . Another strategy would be to use the conditional expectation on action space:

$$s^t = \pi_{\text{avg}}(o^t) = \sum_s s \cdot q(s|o^t, T = 1)$$

Note that this conditional expectation is different from that that would operate in cartesian coordinates. In particular, using a log-polar action map provides with an intrinsic prior for saccades closer to the fixation point (see Figure 2).

2.5 Higher level inference: choosing the processing pathway

Inferring the target location and identity sums up in our case to select a saccade in order to infer the target category from the future visual field. It is likely, however, that a saccade may not provide the expected visual data, and that a corrective saccade may be needed to improve the visual recognition. More generally, choosing to move the eye or to issue a categorical response from the available data resorts to select one processing pathway over the other: either realize the saccade or guess the category from the current foveal data. In order to make this choice, one must guess whether the chance of success is higher in the present, given the current visual field, or in the future, after the next saccade.

This, again, can be expressed under the active inference setup. Let $p(T|f(x_0))$ the probability of success when processing the foveal data, as provided by the “What” network. Under the policy π (provided by the “Where” network), the decision decomposes into a binary choice between issuing a saccade or not. This decision should rely on comparing $p(T|f(\rho(\pi(o)))$ (the future accuracy) and $p(T|f(x_0))$ (the current accuracy). The active inference comes down here to a binary choice between actuating a saccade or “actuating” (testing) the categorical response.

Interestingly, the log difference of the two probabilities

$$\log p(T|f^t(\rho(\pi(o^t))) - \log p(T|f^t(x_0)) \sim \log A^t(\rho(\pi(o^t))) - \log A^t(x_0) \quad (1)$$

can be seen as an estimator of the *information gain* provided by the saccade. Choosing to actuate a saccade is thus equivalent to maximising the information gain provided by the new visual data, consistently with the classic “Bayesian surprise” metric [IB09]. Expanding over purely phenomenological models, our model finally provides a biologically interpretation of the information gain metric as a high-level decision criterion, linked to the comparison of the output of the two principal visual processing pathways.

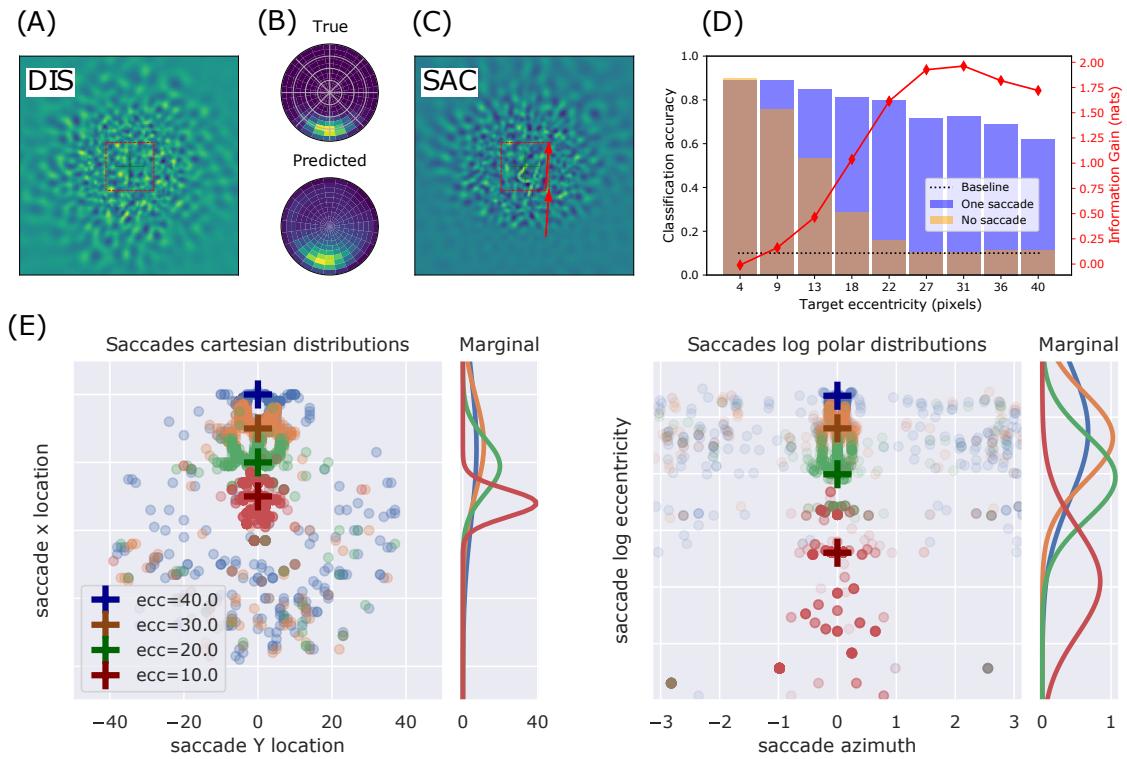


Fig. 2. Example of active vision after training the “Where” network. Digit contrast set to 70%. From left to right: (A) Magnified reconstruction of the visual input, as it shows off from the primary visual features through an inverse log-polar transform. (B) Color-coded radial representation of the output accuracy maps, with dark blue for the lower accuracies, and yellow for the higher accuracies. The network output (“Predicted”) is visually compared with the ground truth (“True”). (C) Visual field shift obtained after doing a saccade. (D) The final classification rate is plotted in function of the target eccentricity. The transparent orange corresponds to the pre-saccadic accuracy from the central classifier (‘no saccade’). The blue bars correspond to the post-saccadic accuracy (‘one saccade’), averaged over 1000 trials per eccentricity. Red line : empirical information gain, estimated from the accuracy difference. (E) Saccades distribution for different target eccentricities. The same saccades are plotted in (pixel) Cartesian coordinates on the left, and in log-polar coordinates on the right. The Cartesian coordinates correspond to the effector space while the log-polar coordinates correspond to the motor control space. In both cases, the empirical marginal distributions over one axis are shown on the right side.

3 Results

After training, the “Where” pathway can predict an accuracy map, whose maximal argument drives the eye toward a new viewpoint with a single saccade. There, a central snippet is extracted, that is processed through the “What” pathway, allowing to predict the digit’s label. Saccades distributions and classification success statistics resulting from this simple sequence are presented in Figure 2. The digit contrast parameter is set to 70% and the eccentricity varies between 4 and 40 pixels. 1000 saccades are sampled for different series of input visual fields containing a target with a fixed eccentricity, but a variable identity, a variable azimuth and a variable background clutter. The full scripts for reproducing the figures and explore the results to the full range of parameters is available at <https://github.com/laurentperrinet/WhereIsMyMNIST>.

Figure 2A-C provides an example of our active visual processing setup. The initial visual field (Fig. 2A) is processed through the “Where” pathway, providing a predicted accuracy map (compared with the true accuracy map in Fig. 2B)). The maximal argument of the accuracy map allows to actuate a saccade. The resulting visual field is provided in Fig. 2C, and the classification is done on the central part of the visual field only (red square). The empirical classification accuracies are provided in Figure 2D, for different eccentricities. They are averaged over 1,000 trials both on the initial central snippet and the final central snippet (that is, at the landing of the saccade). The (transparent) orange bars provide the initial classification rate (without saccade) and the blue bars provide the final classification rate (after saccade). As expected, the accuracy decreases in both cases with the eccentricity, for the targets become less and less visible in the periphery. The decrease is rapid in the pre-saccadic case: the accuracy drops to the baseline level for a target distance of approximately 20 pixels from the center of gaze, consistent with the size of the target. The post-saccadic accuracy provides a much wider recognition range, with a slow decrease from about 90% recognition rate up to about 60% recognition when the target is put at 40 pixels away from the center. An estimate of the information gain provided is provided through a direct comparison of the empirical accuracies (red line). Here an optimal information gain is obtained in the 25-35 eccentricity range.

The lower accuracy observed at larger ranges is an effect of the visual signal bandwidth reduction at the larger eccentricities, that do not allow to accurately separate the target from the background. The spatial spreading of the saccades obtained at different eccentricities is represented on Figure 2E. The same saccades have been represented in Cartesian (pixel) coordinates (left figure) and in log-polar coordinates (right figure). By construction, the log-polar processing, implemented in the “Where” visuo-spatial pathway, leads to a decrease in saccade precision with respect to the eccentricity. This decreasing precision is illustrated by the higher variance of the saccades distribution observed at higher eccentricities, in the Cartesian space of the saccade realization. Interestingly, the variance of the marginal distribution of the saccades along the eccentricity axis is close to constant when represented in the log-polar space (right figure), that is the space of the (collicular) motor command. From 10 to 30 pixels away from the center, the precision of the command is invariant with respect to the eccentricity. The lower precision observed at about 40 pixels eccentricity only reflects a lower detection rate. Due to the log-polar construction of the motor map, the motor command (falsely) appears to display the same precision at various eccentricities. As it would be the case with a more detailed model of the motor noise, this log-polar organization of the control space can be interpreted as a natural re-normalization, helping to counteract the precision loading that would otherwise be attached with the larger saccades, helping to provide a more uniform spread of the motor command in the effector space.

4 Discussion and perspectives

We proposed a computer-based framework allowing to implement visual search under bio-realistic constraints, using a foveated retina and a log-polar visuo-motor control map. A simple “Naïve Bayes” assumption justifies the separation of the processing in two pathway, the “What” visuo-semantic pathway and the “Where” visuo-spatial pathway. The predicted classification rate (or classification accuracy), serves as a guiding principle throughout the paper. It provides a way to link and compare the output of both pathways, serving either to select a saccade, in order to improve the chance of success, or to test a categorical response on the current visual data.

Future work should explore the application of this architecture to more complex tasks, and in particular to a more ecological virtual experiment consisting in classifying natural images. Finally, we used here the log-polar retinotopic mapping as a constraint originating from the anatomy of the visual pathways. At the temporal scale of natural selection, one could also consider this mapping as the emergence of an optimal solution considering an ecological niche, explaining for instance why foveal regions are more concentrated in predators than in preys.

References

1. Connolly, M., and Van Essen, D.: The representation of the visual field in parvcellular and magnocellular layers of the lateral geniculate nucleus in the macaque monkey. *J of Comp Neurology* 226(4) (1984)
2. Daucé, E.: Active fovea-based vision through computationally-effective model-based prediction. *Frontiers in Neurorobotics* 12 (2018). doi: [10/gfrhgj](https://doi.org/10/gfrhgj)
3. Eckstein, M.P.: Visual search: A Retrospective. *Journal of Vision* 11(5), 14–14 (2011). doi: [10/fx9zd9](https://doi.org/10/fx9zd9)
4. Fischer, S., Sroubek, F., Perrinet, L., Redondo, R., and Cristobal, G.: Self-invertible 2D log-Gabor wavelets. *International Journal of Computer Vision* 75(2) (2007). doi: [10.1007/s11263-006-0026-8](https://doi.org/10.1007/s11263-006-0026-8)
5. Friston, K.: The free-energy principle: a unified brain theory? *Nature reviews neuroscience* 11(2) (2010)
6. Friston, K.J., Adams, R.A., Perrinet, L., and Breakspear, M.: Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology* 3 (2012). doi: [10.3389/fpsyg.2012.00151](https://doi.org/10.3389/fpsyg.2012.00151)
7. Itti, L., and Baldi, P.: Bayesian surprise attracts human attention. *Vision Research* 49(10), 1295–1306 (2009). doi: [10.1016/j.visres.2008.09.007](https://doi.org/10.1016/j.visres.2008.09.007)
8. Javier Traver, V., and Bernardino, A.: A review of log-polar imaging for visual perception in robotics. *Robotics and Autonomous Systems* 58(4), 378–398 (2010)
9. Joel, D., Niv, Y., and Ruppin, E.: Actor-critic models of the basal ganglia: New anatomical and computational perspectives. *Neural networks* 15(4-6), 535–547 (2002)
10. Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), 2278–2324 (1998). doi: [10/d89c25](https://doi.org/10/d89c25)
11. Perrinet, L.U., Adams, R.A., and Friston, K.J.: Active inference, eye movements and oculomotor delays. *Biological Cybernetics* 108(6), 777–801 (2014). doi: [10.1007/s00422-014-0620-8](https://doi.org/10.1007/s00422-014-0620-8)
12. Sanz-Leon, P., Vanzetta, I., Masson, G., and Perrinet, L.: Motion Clouds: Model-based stimulus synthesis of natural-like random textures. *J of Neurophys* 107(11) (2012). doi: [10.1152/jn.00737.2011](https://doi.org/10.1152/jn.00737.2011)
13. Sparks, D.L., and Nelson, I.S.: Sensory and motor maps in the mammalian superior colliculus. *Trends in Neurosciences* 10(8), 312–317 (1987)
14. Watson, A.B.: A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of vision* 14(7), 15–15 (2014)