Evolving an Artificial Dorsal Stream on Purpose for Visual Attention

León Dozal, Gustavo Olague*, and Eddie Clemente

Abstract. Visual attention is a natural process performed by the brain, whose functionality is to perceive salient visual features, and which is necessary since it is impossible to focus your sight at two things during the same indivisible time. This work is devoted to the task of evolving visual attention programs through organic genetic programming. The idea is to state the problem of visual attention, which is normally divided in two parts: bottom-up and top-down, in terms of a unique approach based on a teleological framework. Indeed, this paper explains how visual attention could be understood as a single mechanism that is designed according to a given purpose. In this way, genetic programming is used to design top-notch visual attention programs. Experimental results show that this new approach can contrive solutions useful in the solution of "top-down and bottom-up" visual attention problems. In particular, we present a solution to the size popout problem that was unsolved previously in the literature.

1 Introduction

The brain can be extremely complex and despite rapid scientific progress much about how the brain works remains a mystery. In nature, there is a large diversity of brain anatomies that are characterized by the specialization of visual systems. Such diversity shows the power of evolution through adaptation. In this way, it has been argued that the evolution of specific visual mechanisms in the primate brain is the

León Dozal · Gustavo Olague
CICESE, Carretera Ensenada-Tijuana No.3918, Zona Playitas, Ensenada, B.C., México
e-mail: leon.dozal@gmail.com, olague@cicese.mx

Eddie Clemente
Tecnológico de Estudios Superiores de Ecatepec, Avenida Tecnológico S/N,
Esq. Av. Carlos Hank González, Valle de Anáhuac, Ecatepec de Morelos
e-mail: eddie.clemente@gmail.com

* Corresponding author.

O. Schütze et al. (Eds.): EVOLVE – A Bridge between Probability, AISC 175, pp. 371–385.
springerlink.com

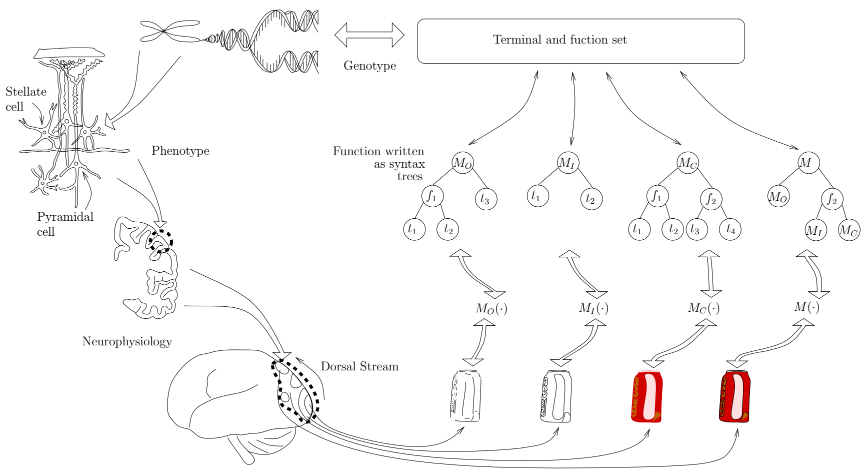


Fig. 1 This figure illustrates the analogy between the natural and artificial systems. The idea is based on replicating the functionality of a set of artificial tissues that conform what we called the organic genetic programming (OGP).

product of natural selection [2]. Contrary, in the past it was widely believed that human observers construct a complete representation of everything in their visual field [6, 21]. This has been amply refuted by a large amount of research. The visual attention is without a doubt one of the most important mechanisms in the visual system because the brain, or visual cortical areas, are unable to process all information received along the entire visual field. In this way, there are two basic process that define the problem of visual attention. The first basic phenomenon is the limited capacity of information processing. At one given time, only a small amount of information available to the retina can be processed and used for control of some specific behavior. The second basic phenomenon is selectivity; in other words, the ability to filter unwanted information [5].

In this work, we follow the idea that visual attention is controled by both cognitive, or top-down (TD) factors, such as knowledge, expectation, and current goals; as well as, reactive stimulus, or bottom-up (BU) factors, that refers to sensory stimulation like gaze, focus, and cuing, see [4]. Moreover, the low level mechanisms for feature extraction act in parallel over the entire visual field using the TD and BU systems in order to provide the signs that highlight the image regions. Afterwards, attention is focused sequentially on the highlighted regions of the image in order to analyze them in more detail [22, 12].

1.1 Problem Statement

As was mentioned previously selectivity is a quality of the visual attention process. Today, many researchers believe that it is necessary to implement this property within artificial systems. The answers to the questions: what features should it use? and when to use those features? is not evident. Moreover, a problem arise after the feature detection stage known as feature combination. Combining different features, such as color, orientation and shape, within a single saliency representation becomes complex since these features came from different visual dimensions. The complexity increases when you are looking for a particular object and it is necessary to filter the information to stress the features of the desired object. In this work organic genetic programming (OGP) is used to address this problem. In this way, the idea is to apply OGP as the mechanism to obtain the most suitable visual attention programs (VAPs) that are capable of pursuing different goals.

2 Visual Attention Processing

This section proposes a new approach for visual attention with the aim of organizing the whole system as a single functional entity that changes its operation according to a purpose, but without changing its structure. In this way, contrary to most traditional approaches that represent the visual attention model through the division of the process into reactive and volitive parts, our proposal provides the simplicity and uniqueness to endow a machine with the ability of designing visual attention programs. Next, the main works are reviewed in order to understand our definition of visual attention.

In the early 80's the feature-integration theory for visual attention was proposed by Treisman and Gelade[22]. Actually, this theory is considered as the most widely accepted paradigm for visual attention within the cognitive sciences community, and it is used as the fundamental computational model for "bottom-up" visual attention. In a first stage, the feature-integration theory proposes that the features of the whole visual scene are perceived in parallel. Next, in a second stage, all features are detected through the stimuli and are integrated into a coherent representation. In a third stage, the stimuli are processed serially by focusing the visual attention on them. In this way, when visual attention is fixed on a particular stimulus, the characteristics, around the attended area, are merged to form a single object. Therefore, it is said that visual attention should serve as a "glue" that combines the features of an individual object to obtain a unique and coherent representation.

2.1 Classical Approach to Visual Attention

The visual attention functionality, regarding the localization of objects, is related to the brain areas around the dorsal stream. Thus, the dorsal pathway is defined as projecting from V1 through V2, V3, middle temporal area (MT), medial superior temporal area (MST) and finally to the posterior parietal cortex, see [23]. Nevertheless, there is a lack of consensus about the specific brain areas, structure and functionality, that conform the dorsal stream. For example, in another theory the dorsal stream is also known as the "how" stream [15]; while, in the work described in [1] the dorsal stream areas do not correspond to the literature.

Nowadays, classical explanations of visual attention are in agreement that the dorsal stream could be seen as a theory that is influenced by BU and TD factors. In this way, it is affirmed by [4] that there are two interacting neural systems involved in the control of BU an TD factors that control visual attention. As a result, the dichotomy of visual attention has inspired several computational models that are commonly based on only one of these two factors. For example, the research in computational neuroscience has traditionally separated their study; as well as, the implementation of visual attention using a benchmark system for human-visual gaze estimation [17, 3] or for the solution of object recognition tasks. Contrary to this line of research, we propose to study visual attention from a teleological standpoint as a way of unifying through this framework both factors with the intention of considering visual attention as a single mechanism.

Next, both BU and TD factors are reviewed in order to introduce our approach with the aim of understanding the structure and functionality of visual attention. In this way, both factors should be studied through a unique process that is capable of adapting itself according to the pursued goal or goals.

2.1.1 "Bottom-Up" Control for Visual Attention

In the literature the idea of BU visual attention is related with involuntary attention, which is usually compared with the concept of a spotlight. This metaphor has been used by Posner *et al.* [18] to explain that visual attention operates "as a spotlight which improves the detection of events in their proximity".

Actually, one of the best and easiest ways of implementing a set of tests is to study BU attention in terms of visual search. Commonly, the exploratory task is studied experimentally using a set of images containing challenging visual stimuli that are presented to an observer. For each image there is an object called target that is different from the rest. Today, the existing computational models are mostly bottom-up models based on the feature-integration theory [22]. The first biologically, neurologically, and plausible computational model for BU visual attention was proposed by Koch and Ullman [12]. Later, Milanese [14] proposed a visual attention system that uses mechanisms, inspired from biological processes, which were adopted by the research community to create a whole new trend in visual attention systems. Some of these processes are the color opponencies such as: red-green and blue-yellow; as well as, the center-surround difference present in the receptive field of the cortical cells. One of the most well-known models is probably that of Itti *et al.* [10], which provided a software that popularize these theoretical processes. In summary, this can be considered as a very detailed model that proposes simple solutions to complex issues. Thereafter, another breakthrough was proposed by Rensink [20, 19] who introduced the notion of proto-objects and the interpretation of the apparent blindness of observers to recognize dramatic changes within a scene. Finally, Walter and Koch [24], showed that the proposed model can enhance the task of object recognition through the application of the concept of proto-object for visual attention tasks.

2.1.2 "Top-Down" Control for Visual Attention

Today, there is an agreement that TD cues play a key role in the processing of visual information. In particular, it is known that there are numerous connections between higher and simpler information processing areas. In this way, it is said that voluntary attention takes more time and effort to accomplish compared to involuntary attention. This is because the target shares with the distractors two or more features, which forces the observer to perform a scanning of the whole scene.

The TD visual phenomenon, just explained, is studied in psychophysics; in fact, TD factors are usually investigated through the so-called "cuing experiments". This type of experiments consists in presenting a "cue" that guides the observer's attention toward the target. In this way, it is said that cues may indicate *where* is the target, like in the case of an arrow pointing towards the target, or by answering the question of *what* is the target by means of finding the similarities between a picture, or written description of the target, see [7]. Thus, there have been several attempts to implement models using TD cues. For example, Oliva *et al.* propose an attentional model that uses knowledge about the distribution of features over the image in order to select salient regions [16]. Peters and Itti [17] proposed a combined model BU/TD, in which they measure the ability of the model to predict the saccades of people playing video games. In this way, they improved the prediction by a margin that doubles the performance obtained by the BU model. The TD part computes a feature vector describing the "gist" of the image with the positions of saccades obtained from real observers that are used to train the model. Finally, a feature vector is calculated to generate the saccades prediction map. Recently, Borji, *et al.* [3] follow the same line of research proposed by Peters and Itti, but the system is based on a different approach that determines the position of the saccades with respect to the observer by applying a set of robust classifiers.

Thus, from a computational modeling standpoint TD factors are not a trivial task; in other words, emotions and desires are difficult concepts to model; hence, such kind of cognitive concepts are still challenging within computer science. Nevertheless, from a pragmatic point of view, there are goals to achieve. Therefore, a purpose should not be confused with a desire; when we refer to a purpose, we talk in terms of whether the goals are achieved or not. Here lies the importance of modeling TD and BU mechanisms in teleological terms.

2.2 An Unified Approach of Visual Attention

Aristotle defined the final cause or *telos* as that for which something is done, its purpose. Also distinguishes between the *telos* and desire, consciousness and intelligence. Therefore, according to Aristotle, an organism like a seed has a purpose just as a person. Latter, Kant [11] wrote, in the "Analytic of Teleological Judgment", that organisms must be regarded in teleological terms, and in the "Dialectic of Teleological Judgment", he attempts to reconcile this teleological conception of organisms with a mechanistic account of nature. Everything can be completely explained by causality, except the organisms.

From our standpoint, attention is the result of a single mechanism that is designed to obey a general purpose; as well as, different particular purposes. For example, the most primitive purpose for life could be survivorship. But, the achievement of survivorship depends on many other particular purposes; for example, prey hunting, mating, predator escape, etc. In this sense, visual attention is capable to adapt itself to the kind of goal that depend on the current purpose of the organism. In order to accomplish such task, it is necessary a unique and general visual attention structure capable of performing, by some temporal readaptation, the necessary functions to achieve that aim or intention. Furthermore, considering the fact that most of the tasks involved in the design of BU and TD factors are complex, we could say that the space of possible readaptations is at least very large and discrete. Therefore, we defined visual attention as follows.

Definition 1 (Visual Attention) *Visual attention is a process that designs a relationship between the different properties of the scene, which are perceived through the visual system with the aim of selecting a particular aspect.*

For these reasons, we consider the visual attention as a single computational structure that performs BU and TD processes. In consequence, in this work visual attention is studied within a unified framework in order to evolve visual attention programs (VAPs) that will be adapted for specific purposes.

3 Purposive Evolution for Visual Attention

The theory of evolution is not exempt of the concept of purpose. Charles Darwin was the one who brings the concept of purpose into consideration. Note that Darwin uses the term final cause systematically in his writings as it is documented by Lennox [13]. On the other hand, Barton [2] explains the evolution of primates brains in terms of the specialization of visual mechanisms; such as visual attention. Thus, this section describes the general structure of attention, which is biologically inspired and will be evolved to suit different purposes. The resulting evolved programs will be known as visual attention programs (VAPs). Moreover, following the same direction of Treisman, the description of the general approach is divided into two main stages: acquisition and integration.

3.1 Acquisition of Early Visual Features

In previous works of artificial visual attention, the operators are established according to the knowledge in neuroscience. Moreover, it is widely recognized that the operation of the visual cortex, specifically the dorsal stream, is a product of the evolutionary process. In this way, we propose to use evolutionary computation to obtain these artificial visual operators. In summary, this work explains how to use specialized evolved visual operators (EVOs) for the acquisition of visual dimensions such as color, orientation and shape. Next, the EVO features used within the VAP are defined.

Table 1 Functions and terminals used by EVO_O to create the orientation visual map VM_O

$F_O = \{+,\ -,\ \times,\ \div,\ \|+\|,\ \|-\|,\ \sqrt{I_{T_O}},\ I_{T_O}^2,\ log_2(I_{T_O}),$ $G_{\sigma=1}, G_{\sigma=2},\ \|I_{T_O}\|,\ \frac{I_{T_O}}{2},\ D_x,\ D_y\}$
$T_O = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v, G_{\sigma=1}(I_r), G_{\sigma=2}(I_r),$ $D_x(I_r), D_y(I_r), D_{xx}(I_r), D_{yy}(I_r), D_{xy}(I_r), \ldots\}$

3.1.1 Orientation

In previous work the characteristic of orientation for images was only computed in gray scales. Thus, our work proposes to evolve the property of orientation along the different color bands of the image. In this way, a rich set of information is generated because the edges, corners, and other similar features could appear more highlighted with the color bands. Therefore, the evolutionary approach evolves a function $EVO_O : I_{color} \rightarrow VM_O$ that cooperates with the VAP in order to accomplish a purpose. The resulting EVO_O operation is a visual map VM_O for which the pixel value represents the feature prominence; in such a way, that the larger the pixel value, the greater the orientation prominence of the feature. This computation is performed through a set of functions and terminals that are provided in Table 1. The notation that was used is as follows. I_{T_O} can be any of the terminals in T_O; as well as, the output of any of the functions in F_O; D_u symbolizes the image derivatives along direction $u \in \{x, y, xx, yy, xy\}$; G_σ are Gaussian smoothing filters with $\sigma = 1$ or 2.

3.1.2 Color

In biology, the color is encoded through photoreceptor cells known as cones, which are located in the retina. However, a special case is the yellow color which is not perceived in the cones but in the retinal ganglion cells. Then, the dorsal pathway is composed of several tissues V1, V2 and V4, whose cells respond to color features. In this work the characteristics of color information that will be used as the building blocks to construct the *VAPs* are color opponencies and simple arithmetic operations between the different color bands in the corresponding color space. In the same way, as in EVO_O, the evolutionary process uses a set of functions and terminals provided in Table 2 to evolve the feature in the color space. The result is a visual map $EVO_C : I_{color} \rightarrow VM_C$ containing the color prominent features.

Table 2 Functions and terminals used by EVO_C to create the color visual map VM_C

$F_C = \{+,\ -,\ \times,\ \div,\ \|+\|,\ \|-\|,\ \sqrt{I_{T_C}},\ I_{T_C}^2,\ log_2(I_{T_C}),$ $Exp(I_{T_C}, Complement(I_{T_C})\ \}$
$T_C = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v, RG_{oppn}, YB_{oppn}\ \}$

3.1.3 Shape

As in previous dimensions, the evolutionary process uses a set of functions and terminals provided in Table 3 to characterize the shape information. Note that we propose to describe these features through mathematical morphology. The result is a visual map $EVO_S : I_{color} \rightarrow VM_S$ containing the shape prominent features. This part is evolved with genetic programming with the aim to provide the information about shape and structure of the object of interest within the image. We would like to remark that the application of this kind of morphological functions has not been applied in previous research studying the ventral and dorsal streams.

Table 3 Set of functions and terminals used by EVO_S to create the shape visual map VM_S

$F_S = \{+, -, \times, \div, round(I_{T_S}), \lfloor I_{T_S} \rfloor, \lceil I_{T_S} \rceil, dilation_{diamond}(I_{T_S}), dilation_{square}(I_{T_S}), dilation_{disk}(I_{T_S}), erosion_{diamond}(I_{T_S}), erosion_{square}(I_{T_S}), erosion_{disk}(I_{T_S}), skeleton(I_{T_S}), boundary(I_{T_S}), hit - miss_{diamond}(I_{T_S}), hit - miss_{square}(I_{T_S}), hit - miss_{disk}(I_{T_S}), top - hat(I_{T_S}), bottom - hat(I_{T_S}), open(I_{T_S}), close(I_{T_S}) \}$
$T_S = \{I_r, I_g, I_b, I_c, I_m, I_y, I_k, I_h, I_s, I_v\}$

Finally, to obtain the intensity of pixels in the image the model averages the red, green and blue values for each pixel. The result of this operation is a visual map VM_I in which the pixel represents the prominence over the intensity space.

3.1.4 Computing the Conspicuity Maps

The conspicuity maps (CMs) are obtained by means of a center-surround function that is applied in order to simulate the center-surround receptive fields. This natural structure allows the ganglion cells to measure the differences between firing rates in center (c) and surroundings (s) areas of ganglion cells. At this stage, we have one *CM* for each feature. The *CM* is obtained as proposed in the Walther and Koch model [24]. Finally, the *CMs* are combined to obtain a single saliency map as explained in the next section.

3.2 Feature-Integration for Visual Attention

The saliency map (SM) defines the place for the most prominent locations of the image; given the characteristics of intensity, orientation, color and shape. In other words, the objective of this stage is to decide where attention could be directed at any given time. In this work, the problem statement considers that the task must be

Table 4 Set of functions and terminals used by EFI to create the object saliency map OSM

$F_{fi} = \{+, -, \times, \div, \|+\|, \|-\|, \sqrt{I_{T_{fi}}}, I^2_{T_{fi}}, Exp(I_{fi}), G_{\sigma=1}, G_{\sigma=2}, \|I_{T_{fi}}\|, D_x, D_y\}$
$T_{fi} = \{CM_I, CM_O, CM_C, D_x(CM_I), D_y(CM_I), D_{xx}(CM_I), D_{yy}(CM_I), D_{xy}(CM_I), ...\}$

addressed to achieve a specific goal. As a result, if the task needs to accomplish a purpose; then, the main criterion should be the one that guides the suitable combination of characteristics. Therefore, we decided to evolve the integration of *CMs* through a function that we called Evolved Feature Integration (EFI). Therefore, the *VAPs* provide a dynamic structure since the *EVOs* can be selected using a fusion process executed by the EFI. This process considers different combinations of *CMs* to complete the entire process. Once the integration of features is performed, we get an optimized saliency map (OSM) indicating the location of the most prominent regions within the original image, known as proto-objects (P_t). The definition of the *EFI* function is as follows:

$$EFI : CM_l \rightarrow OSM \quad ; \quad l \in \{O, C, I\}$$

The evolutionary method uses the set of functions and terminals, listed in Table 4, to create a fusion operator that highlights the features of the object of interest.

4 Organic Genetic Programming

In this section, we describe the main aspects for the evolution of *VAPs* using the organic genetic programming (OGP) strategy. In the OGP the chromosome is composed of several genes that are represented each one with a tree structure. At the gene level the genetic operations are performed like in the classical genetic programming. While at the chromosome level the whole genotype is described by the parallel set of functions acting over the color, shape, and orientation dimensions. The design of the OGP embody an organic motivation in a sense of describing an organ or tissue, as a part of a living organism, and their complexity. We introduce a set of new concepts in order to deal with the evolution of complex structures, which are explained below.

In the experiments the OGP goal is to discover a program that learns to attend a prominent object using a set of training images. In this work, the *VAP*'s genotype is considered as robust because it is capable of encoding the phenotype of an artificial dorsal stream. In other words a genotype consisting of three to four trees is composed of different and specialized operations. Thus, each tree has its own independent set of functions and terminals, which are listed in Tables 1, 2, 3 and 4; according to orientation, color, shape and feature integration respectively.

The first one encodes the orientation feature similar to the orientation-sensitive cells of V1 [9]. The second one represents the color feature in an analogy to the photo-receptor cells presented in the retina; as well as, the color sensitive cells of the layers V1 and V4 of the visual cortex. The third one models the shape feature that characterizes shape-sensitive cells present in layers V2 and V4 of the brain. Finally, the fourth one encodes the way in which the features are combined to obtain the saliency map, or operation of the posterior parietal cortex [8].

5 Experiments and Results

The following experiments are divided in two parts, according to the goal that the *VAP* is attempting to reach. The OGP is basically the same for both experiments, the only parts that change are the fitness function, which encodes the purpose, and the set of images utilized for training. The fitness function of the OGP is the characterization of the purpose, the answer to the "what are the individuals for?". In other words, it is the way in which the purpose is implemented as a computer programming.

5.1 Evolution of VAPs for Aiming Scene Novelty

The first set of experiments are stated in terms of *visual search*, which is commonly applied like in classical research devoted to visual attention. In this way, the tests are designed to obtain through artificial evolution a VAP that is specially adapted to find the novelty, or asymmetries, in a simple set of images of the kind that are used in psychophysical studies.

5.1.1 Search of Appearance Novelty

The first experiment was conceived with the aim of obtaining a VAP capable of centering attention with respect to appearance novelty. In other words, capable of focusing an object using the shape and the information around the object. Figure 2 shows the $VAP_{triangle}$ that was obtained by the OGP. We can remark that the $VAP_{triangle}$ utilizes only the color dimension. The proposed solution is to regularize the image through the logarithm function. This process reduces the contrast between the black and white areas, and as a result, the regions around the triangle are highlighted after the central-surround processing and evolved feature integration steps. Thus, the functions obtained by the OGP are listed below:

$$
\begin{aligned}
EVO_O &:= \quad D_y(I_G) \\
EVO_C &:= \quad log(log(I_B)) \\
EFI &:= \quad \|D_x(CM_C)\|
\end{aligned}
$$

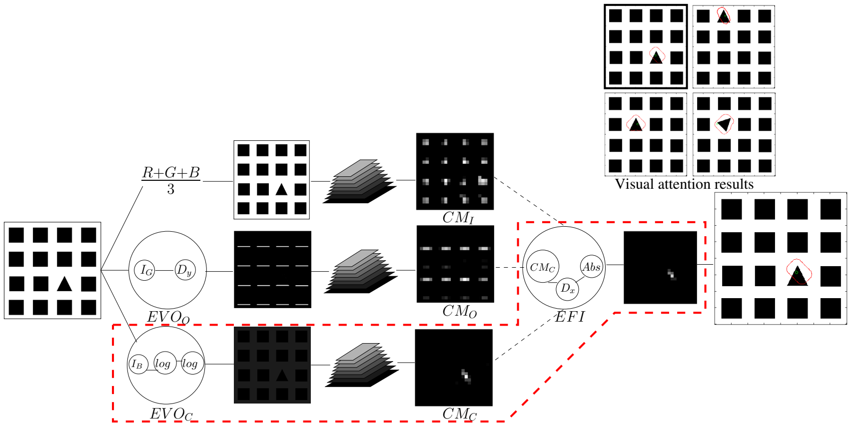


Fig. 2 Bottom-Up image testing of novelty. This figure depicts the best visual attention program that was evolved with organic genetic programming to attend the triangle.

The evolution used only one image per training, This image shows the highlighted black square at the top-right corner of Figure 2. The extra images illustrate the results achieved during a set of tests considering rotation and translation; indeed, the triangle was rightly focused.

5.1.2 Search of Size Novelty

The experiment described next is important because, according to the literature, it has not been solved previously by any computational method devoted to the solution of visual attention. The main reason may be due to the overlook of the characteristic of size, and the lack of a suitable choice of functions within the problem statement. Thus, in order to solve this problem for this particular experiment, see Figure 3, it was decided to increase as an extra dimension the property of shape (EVO_S), which is computed as explained in section 3.1.3 using the fundamental operations of mathematical morphology. Hence, the best set of functions obtained by the OGP are listed below:

$$
\begin{aligned}
EVO_O &:= D_x D_y D_{xx}(Y) + tresh\left(\frac{D_x D_y D_{xx}(Y)+0.93}{0.83}\right) \\
EVO_C &:= I_V - I_R \\
EVO_S &:= ((I_G + 0.90) \oplus Sqr) \oplus Sqr \\
EFI &:= \left| D_y(CM_S) - \frac{CM_C}{\left|CM_O - \frac{CM_C}{D_y(CM_O)}\right|} \right|
\end{aligned}
$$

where Sqr denotes a square structuring element over the dilation operator $\oplus$.

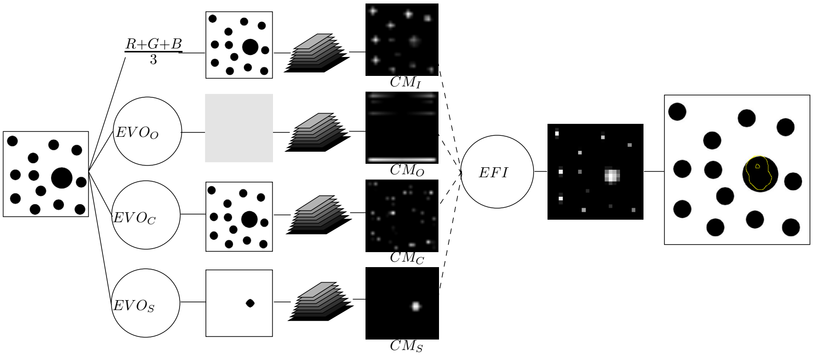


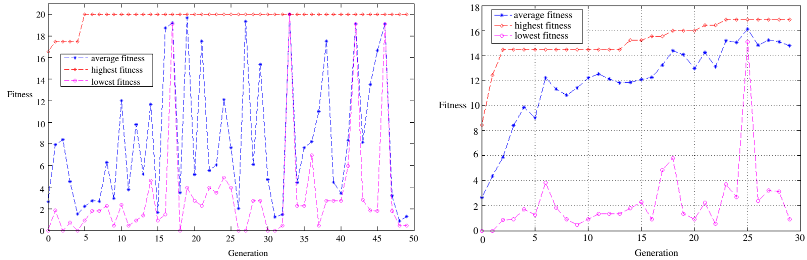
Fig. 3 Bottom-Up testing for the size popout problem. This figure provides an example of a pop-out effect, the big circle, that previous visual attention programs were unable to detect, see [7]. Indeed, the evolved visual attention program was able to detect the saliency in the image related to the single big object.

5.2 *Evolution of VAPs for Aiming Specific Targets*

In this section, the obtained *VAPs* and their performance are presented for the case of TD tasks. Figure 4 provides the statistics of the top-down runs plotting the average, highest, and lowest values. The examples illustrate that for each target object: red can and traffic signals, a solution could be attained without changing the proposed computational framework. During the training stage, the FOA_{coke} detected the object of interest, in this case the coke, with a successful rate of 100% after using 44 images. In this way, during the testing stage the FOA_{coke} detected the object of interest with a rate of 88.13% using 59 images. In this way, from the 59 test images the coke was detected in 52 occasions. Moreover, the percentage of detection increases after considering a second attempt since the coke was detected in one additional image; producing a total of 53 images that represent 89.83% of the total, see Figure 5. This best individual brought into consideration the reflectance, which is a feature that in some images of the can is useful for the solution of the problem. Next, the functions obtained by the OGP are listed.

$$EVO_O := \quad Dx(I_K) - Dy(I_Y)$$
$$EVO_C := \quad (Exp(I_G)/(I_H)^2)^2$$
$$EFI := \quad Dx(CM_C)$$

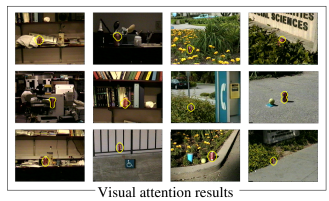
As a final experiment, the training stage was applied to the FOA_{signal} in order to detect the object of interest, in this case the traffic signal, with a percentage rate of 88.89%; in other words, in 40 of the 45 training images. Next, during the testing stage the FOA_{signal} detected the object of interest, traffic signal, in 77.78% of the images. Thus, from 45 testing images the best evolved program correctly detected



(a) Statistics of the red can (b) Statistics of the traffic signal

Fig. 4 These figures show the average, highest, and lowest fitness; for a) the red can run, and b) the traffic signal that produce the respective best individual of Figures 5 and 6.

	Training (N = 44)	Testing (N = 59)	
	P_1	P_1	P_2
VAP_{coke}	100%	88.13%	89.83%



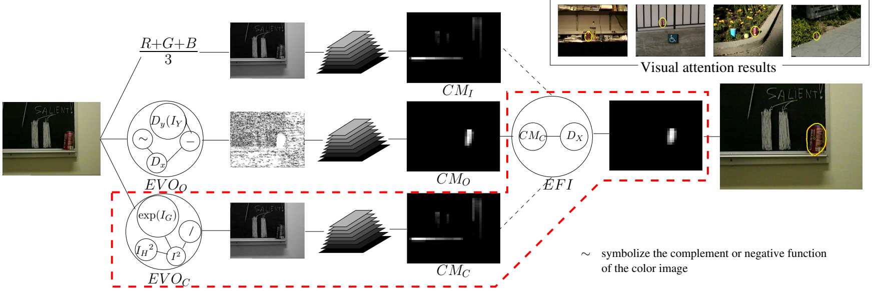


Fig. 5 Top-Down testing for the red can problem. Evolved structure of VAP_{Can} obtained through the OGP to attend the red can in the images

the salient object in 35 images. Moreover, the percentage of detection increases when we consider a second attempt as the signal was detected in five additional images producing a total of 40 images that represent a rate of 88.89%. Finally, in a third attempt the percentage increases to 95.56%, see Figure 6. Next, the best functions that were discovered by the OGP are listed.

$$EVO_O := Half(|G_{\sigma=2}(Half(Half(G_{\sigma=2}(Half(|Dxx(I_M)|))))))|)$$
$$EVO_C := \sqrt[4]{I_B}$$
$$EFI := \left(\sqrt{\sqrt{Dyy(CM_O)} \times CM_O}\right) \times CM_O$$

	Training (N = 45)	Testing (N = 45)		
		P_1	P_2	P_3
$VAP_{traffic}$	**88.89%**	**77.78%**	**88.89%**	**95.56%**

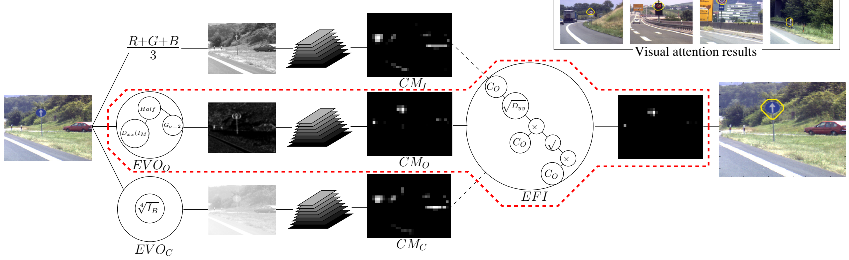


Fig. 6 Top-Down testing for the traffic signal problem. Evolved structure of $VAP_{traffic}$ obtained through the OGP to attend traffic signals in the images

6 Conclusions

This work presents a new and useful approach for understanding visual attention. The experiments are motivated by new ideas about purposive evolution and organic genetic programming. The results confirm that it is possible to obtain VAPs that fulfill the task at hand. Moreover, an original program that solves the size pop-out search task was obtained by our approach, and to our knowledge it is the first time to be achieved. Also, the incorporation of shape dimension, carried out with morphological operations, is an original contribution to the research in visual attention.

References

1. Baluch, F., Itti, L.: Mechanisms of top-down attention. Trends in Neurosciences 34(4), 210–224 (2011)
2. Barton, R.A.: Visual specialization and brain evolution in primates. Proceedings of the Royal Society of London Series B-Biological Sciences 265(1409), 1933–1937 (1998)
3. Borji, A., Sihite, D., Itti, L.: Computational modeling of top-down visual attention in interactive environments. In: Proceedings of the British Machine Vision Conference, pp. 85.1–85.12. BMVA Press (2011)
4. Corbetta, M., Shulman, G.L.: Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience 3(3), 201–215 (2002)
5. Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention. Annual Reviews 18, 193–222 (1995)
6. Feldman, J.A.: Four frames suffice: A provisional model of vision and space. Behavioral and Brain Sciences 8(02), 265–289 (1985)
7. Frintrop, S.: VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search. LNCS (LNAI), vol. 3899. Springer, Heidelberg (2006)
8. Gottlieb, J.: From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron 53(1), 9–16 (2007)

9. Hubel, D.H., Wiesel, T.N.: Receptive fields of single neurons in the cat's striate cortex. Journal of Physiology 148, 574–591 (1959)
10. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20(11), 1254–1259 (1998)
11. Kant, I.: Critique of the power of judgment. Cambridge University Press (2000)
12. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4(4), 219–227 (1985)
13. Lennox, J.G.: Darwing was a teleologist. Biology and Philosophy 8, 409–421 (1993)
14. Milanese, R.: Detecting salient regions in an image: from biological evidence to computer implementation. PhD thesis, Department of Computer Science, University of Genova, Switzerland (December 1993)
15. Milner, D., Goodale, M.A.: The visual brain in action, 2nd edn. Oxford University Press (December 1995)
16. Oliva, A., Torralba, A., Castelhano, M., Henderson, J.: Top-down control of visual attention in object detection. In: International Conference on Image Processing, vol. 1, pp. 253–256 (2003)
17. Peters, R.J., Itti, L.: Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN (June 2007)
18. Posner, M.I., Snyder, C.R., Davidson, B.J.: Attention and the detection of signals. Journal of Experimental Psychology 109(2), 160–174 (1980)
19. Rensink, R.A.: The dynamic representation of scenes. Visual Cognition (2000)
20. Rensink, R.A.: Seeing, sensing and scrutinizing. Vision Research 40, 1469–1487 (2000)
21. Trehub, A.: The Cognitive Brain. MIT Press, Cambridge (1991)
22. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology 12(1), 97–136 (1980)
23. Ungerleider, L.G., Mishkin, M.: Two cortical visual systems. In: Ingle, D.J., Goodale, M.A. (eds.) Analysis of Visual Behavior, pp. 549–585. MIT Press, Cambridge (1982)
24. Walther, D., Koch, C.: Modeling attention to salient proto-objects. Neural Networks 19(9), 1395–1407 (2006)