Un modelo de atención visual para la detección de regiones de interés en imágenes radiológicas

Julio Ernesto Villalón Reina M.D.

Universidad Nacional de Colombia
Facultad de Medicina
Departamento de Imágenes Diagnósticas
Bogotá, Colombia
2014

Un modelo de atención visual para la detección de regiones de interés en imágenes radiológicas

"A computational model of attention for Region-of-Interest selection in radiological images"

Julio Ernesto Villalón Reina M.D.

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magister en Ingeniería Biomédica

Director(a):
Edgar Eduardo Romero Castro M.D., Ph.D.

Línea de Investigación:
Procesamiento de Imágenes Médicas
Grupo de Investigación:
Computer Image and Medical Applications Laboratory (CIMALAB)

Universidad Nacional de Colombia
Facultad de Medicina
Departamento de Imágenes Diagnósticas
Bogotá, Colombia
2014

A mis padres Jorge y Rosalba

Acknowledgments

I would like to thank Dr. Eduardo Romero for his unmatchable wise and tough guidance. Drs. Ricardo Gutiérrez and Francisco Gómez gave invaluable input to this research as members of my committee. Also, this work would not have been possible without the help and materials provided by the staff members and doctors of the Magnetic Resonance Center at the San José Hospital in Bogotá (Colombia), in particular Ricardo Alvarez and Dr. Leonidas Borrero.

Abstract

The detection, segmentation and quantification of multiple sclerosis (MS) lesions on magnetic resonance images (MRI) has been a very active field for the last two decades because of the urge to correlate these measures with the effectiveness of pharmacological treatment. A myriad of methods has been developed and most of these are non specific for the type of lesions, e.g. they do not differentiate between acute and chronic lesions. On the other hand, radiologists are able to distinguish between several stages of the disease on different types of MRI images. The main motivation of the work presented here is to computationally emulate the visual perception of the radiologist by using modeling principles of the neuronal centers along the visual system. By using this approach we were able to successfully detect multiple sclerosis lesions in brain MRI. This type of approach allows us to study and improve the analysis of brain networks by introducing a priori information.

Keywords: Multiple sclerosis, visual attention, artificial vision, magnetic resonance imaging

Resumen

La detección, segmentación y cuantificación de lesiones de esclerosis múltiple (MS) en imágenes de resonancia magnética (MRI) ha sido un área de estudio muy activa en las últimas dos décadas. Esto es debido la necesidad de correlacionar estas medidas con la efectividad de los tratamientos farmacológicos. Muchos métodos han sido desarrollados y la mayoría no son específicos para los diferentes tipos de lesiones, es decir que no pueden distinguir entre lesiones agudas y crónicas. Los médicos radiólogos por su parte son capaces de distinguir entre diferentes niveles de la enfermedad haciendo uso de las imágenes de resonancia magnética de diferentes tipos. La principal motivación de este trabajo es la de emular mediante un modelo computacional la percepción visual del radiólogo, haciendo uso de los principios fisiológicos del sistema visual. De esta manera logramos detectar satisfactoriamente las lesiones de esclerosis múltiple en imágenes de resonancia magnética del cerebro. Este tipo de análisis nos permite estudiar y mejorar el estudio de las redes neuronales al poder introducir información a priori.

Palabras clave: Esclerosis múltiple, atención visual, visión artificial, imágen por resonancia magnética.

Contents

List of Figures

List of Tables

1. Introduction and motivation

Multiple sclerosis (MS) is the most prevalent demyelinating disease in the world. Its prevalence has been estimated to be between 2 and 25 per 100,000 habitants [5]. This large range in prevalence is due to the geographical variance of the disease, i.e. the higher the latitude in the northern hemisphere the larger the prevalence of the disease. In general, demyelinating diseases are characterized by the destruction of the myelin of nerve fibers with relative sparing of axons, nerve cells, and supporting structures. Other pathological hallmarks of this group of disorders is the infiltration of inflammatory cells in a perivascular and particularly paravenous distribution and a distribution of lesions that is primarily in white matter, either in multiple small disseminated foci or in larger foci spreading from one or more centers.
The main symptoms of MS are motor weakness, paraparesis, paresthesias, loss of sight, diplopia, nystagmus, dysarthria, intention tremor, ataxia, impairment of deep sensation, and bladder dysfunction. These symptoms occur most frequently in what is called the relapsing-remitting pattern. This pattern is characterized by initial manifestations that may or may not be noticed by the patient and that improve partially or completely and are then followed after a variable interval by the recurrence of the same abnormalities or the appearance of new ones in other parts of the nervous system.
Pathologically, MS lesions may vary in diameter from less than a millimeter to several centimeters and they are usually localized in the periventricular areas, but only where subependymal veins line the ventricles (body and atria of the lateral ventricles). Other regions that may be affected are the brainstem, spinal cord, and cerebellar peduncles. Although chronic lesions are essentially still demyelinative, partial remyelination has been found to take place on undamaged axons. This finding has attracted the attention as the target for potential pharmacological therapeutics that could potentiate and promote remyelination processes [6].
Magnetic resonance imaging (MRI) has gained a special position for the diagnosis and follow-up of patients with MS. T2 weighted MRI images have the ability to reveal MS plaques in the cerebrum, brainstem, optic nerves, and spinal cord. This plaques are detectable even without any proper MS symptoms. Acute and chronic MS plaques are hyperintense (look white on the image) on T2-weighted spin echo images and even more strikingly obvious on T2-weighted FLAIR (fluid attenuated inversion recovery) images. Especially diagnostic are oval or linear regions of demyelination, oriented perpendicularly to the ventricular surface. This radiological sign is usually referred as “Dawson fingers” because of their characteristic thick elongated appearance. Some of these demyelinating areas may extend into the centrum

semiovale and may reach the convolutional white matter. See figure 1.1 for example MRI images. Morphological changes of the lesions on T2-weighted images across time have been radiologically described, especially the confluence of many small lesions into single big lesions at the poles of the ventricles. The overall trend of T2 lesions is to increase in number and volume over time, a phenomenon also referred to as the "T2 burden of disease". This burden is more severe in the absence of treatment and less so when there has been effective treatment. Thus, the T2 burden of disease has been used as a biomarker in MS treatment trials [7].

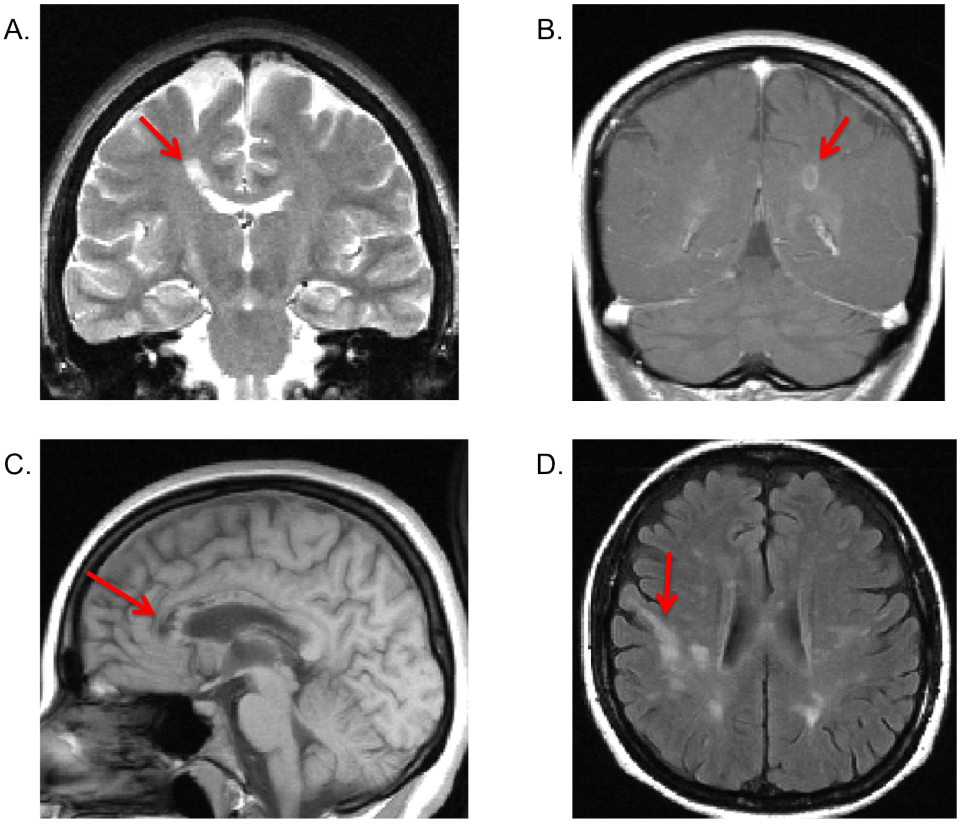

Figure 1.1.: A.) T2 spin echo weighted image showing a typical hyperintense "Dawson finger" starting from the ventricle edge and extending deeper into the white matter. B.) T1 weighted image with contrast medium (Gadolinium) showing a ring enhancement surrounding an acute demyelinating lesion. C.) T1 weighed image showing chronic hypointense demyelinated lesions in the corpus callosum, an important sign of axonal degeneration. D.) T2 FLAIR weighted image with hyperintense lesions in the deep white matter that extent towards the subcortical white matter within the gyri.

Gadolinium enhanced MRI images are used for the detection of acute MS lesions that present early inflammation. These lesions usually last between 4 and 8 months and their detection has been also used as a biomarker for relapsing episodes of MS in drug trials [8]. MRI has been also found to detect plaque remyelination processes. Typically, T1-weighted images

are able to identify the demyelinated areas that progress to axonal damage. These areas are seen in T1-weighted images as "black holes". Axonal loss is an irreversible pathological phenomenon and will ultimately determine the patients outcome. Importantly, Barkhof et al. found that T1 hypointensity within a plaque as well as the magnetization transfer rate were inversely correlated to the degree of remyelination [9], suggesting that MRI is able to differentiate between the pathological stages of the disease.

All of the evidence shown above brings brain imaging into the center of the diagnostic and prognostic approaches for MS, and places challenging and interesting image processing problems to improve and gain the accuracy in detection of demyelinating lesions as well as their size estimation, remyelination and pharmacological treatment effects.

Most of the above mentioned methods rely on the segmentation of the lesions by using the voxel intensity for that purpose. The majority of the algorithms use a multichannel approach, meaning that they combine several types of MRI images, usually the T1-weighted, T2-weighted, PD, FLAIR and contrast enhanced images [10–12]. Although many studies have also used only one modality of MRI image, especially T2-weighted or PD images [13]. Another important distinction is that many methods rely on the manual input of an expert to help the segmentation process to be more accurate, which makes it a semiautomatic method [14], whereas other approaches are absolutely automatic [10]. The other very important distinction to make is the one between supervised and unsupervised methods. The so-called supervised methods rely on prior information. The prior information can be provided as either a brain probabilistic or topological atlas, or as manually pre-segmented and annotated lesions for further classification purposes [15–17]. The unsupervised methods can also be divided into two types, those that classify lesions as outliers based on a previous tissue segmentation of the brain and those that only use the lesion properties to segment them [18, 19].

Although a thorough review of all the algorithms used for the segmentation of MS lesions is beyond the scope of this chapter, it's worth mentioning the most relevant and influential ones. For example, Zijdenbos et al. developed a processing pipeline that denoises the images and thereafter runs a tissue classification algorithm [10]. The algorithm is based on back-propagation artificial neural networks (ANNs). Their system is trained with manually segmented images and the ANN input layer used three different scan modalities (T1, T2 weighted and PD images) and tissue specific probabilistic maps as priors (gray matter, white matter and cerebrospinal fluid probabilistic maps). This is clearly a supervised method with input from a probabilistic atlas and from manually segmented and annotated images. Wells et al. presented an influential unsupervised segmentation algorithm [15]. They first create a Gaussian mixture distribution of intensities to characterize the different brain tissue types as well as the image bias field. By utilizing these tissue properties an expectation maximization algorithm is implemented to adaptively determine the desired "missing" tissue classes (white and gray matter, CSF and MS lesions). They tested their algorithm on a dataset of 1000 T2-weighted images of patients with MS brain lesion. A similar approach has been used by

other authors to quantify the evolution of the burden of disease in MS across time [12, 20]. These researchers also added a connectivity-based partial volume effect correction after the tissue segmentations took place. Warfield et al. also utilized expectation maximization to extract the different tissue types but after they provided prior information from an atlas [16]. They linearly and non-linearly align an atlas (or template) to the subject's space. The atlas has been previously segmented and labeled by hand into cortical and subcortical gray matter structures, white matter and CSF. By using the matched atlas' white matter mask they were able to discriminate between MS lesion vs gray matter and to determine the boundaries of white matter after excluding the CSF. Van Leemput et al. performed a stochastic model-aided segmentation for automatic segmentation of MS lesions [11]. This is also an atlas-supervised segmentation strategy because each voxel was iteratively interleaved into a set of tissue-type models (gray matter, white matter and CSF) and the voxels that weren't explained well by the tissue models were labeled as MS lesions. Udupa et al designed a semiautomatic method based on fuzzy connectedness [14]. The user adds a few starting points within each anatomical brain structure, such as the white matter, gray matter and CSF, and the system then detects each one of them as a fuzzy-connected 3-D object. The MS lesions are detected as holes in the union of these three anatomical objects. A very interesting method for MS lesion segmentation on proton density and T2 weighted images was proposed by Pachai et al. [19]. Their algorithm can be considered as a multi-channel but also an unsupervised method of lesion segmentation. It uses a multi resolution approach, by constructing a Gaussian pyramid of low-pass versions of the original image. Each of these versions is resampled to the original resolution and subtracted from the initial image. This gives a Laplacian pyramid of increasing high-pass representations of the initial image. A local thresholding algorithm estimates the most hyper intense areas that the Laplacian pyramid is able to enhance in the image. After applying morphological tools the external hyperintense CSF areas are left out and the MS lesions are then finally quantified. This method is robust enough to overcome the scanner induced intensity inhomogeneity.

After more than 20 years of research on MS lesion segmentation and quantification, no study has been published yet, to the extend of our knowledge, that uses biologically inspired algorithms to detect demyelinating pathology in brain MRI images. By biologically we mean essentially the use of physiologic principles of the visual system that have been mathematically modeled and that are worth exploring for medical image processing. The main motivation of this work is to bring one of such models to the context of brain imaging so that it is possible to detect MS pathology in MRI images and try to emulate the visual system at high levels of visual expertise like radiology.

When determining the so called "burden of the disease" most of the referred methods try to look for the total extent of the lesions. This is why the majority of the methods use multichannel approaches, since some lesions may appear independently in different types of images. One advantage of visual attention models is that they can be tuned and trained in order to detect and even describe specific types of targets in the scene. When developing

such a model for medical imaging purposes, the aim is to detect and discriminate between different pathological conditions. It is possible to design a system that can accurately detect a particular type of lesion. In this work we hypothesize that by using some basic principles of the visual system it is possible to detect MS lesions. The resulting tool will be specific to detect MS lesions in T2 weighted FLAIR images and can eventually be tuned to describe various types of lesions in MRI.

The combination of the current knowledge about the visual perception of radiologists with existing computational visual attention models may help disentangle many of the challenges concerning problems like accurate MS classification and staging and their relationship to therapeutics. As said before, MRI techniques are rapidly evolving and are already able to distinguish between different pathological stages of brain tissue. In this work we argue that computational visual models may be an efficient way to not only analyze pathological MRI images based on image features (luminance and texture) but also to find a suitable tool to study the neural networks involved in human visual perception.

2. Theoretical background

The training process in radiology implies learning many skills that are necessary to accomplish a reliable diagnosis. This diagnosis is mainly based on the information embedded in the radiological image (X-rays, computer tomography, ultrasound, magnetic resonance imaging), but it also depends on the constantly evolving structured knowledge of the radiology trainee. There are four steps involved in the radiological diagnostic process, that constitute a general framework for problem solving in radiology: searching & detecting - recognizing - deciding. From a behavioral point of view, the first two steps mentioned (searching and detecting) imply the activation of perceptual and cognitive processes specifically related to visual attention. In this chapter, we first give an introduction to some basic morphological and physiological concepts related to the process of vision in humans and primate brains to later connect them to the concept of visual attention, which is the core of the work developed in this manuscript.

2.1. Neurophysiological background of the visual system

An important physiological concept of visual processing is the direction of information flow. Visual information can flow in a bottom-up or a top-down manner through clusters of neurons embedded within the cortex and subcortical grey matter structures (i.e. the lateral geniculate nucleus). Each of these clusters is considered as a stage in the process. Synaptic connections to one visual processing stage are reciprocal, which means that each area receives feed-forward projections from an area earlier in the stream and provides a feedback projection to the same area. The feed-forward pathway provides the bottom-up input to subsequent visual areas. Feed-back projections are thought to be responsible for the top-down modulation.

Another important concept is the segregation of the information into different pathways. In general two main visual processing cortical streams have been described [21]. The ventral or "what" pathway runs from the primary visual area (V1) to the secondary visual area (V2), visual area four (V4), anterior inferior temporal cortex (TE) and posterior inferior temporal cortex (TEO). TE and TEO are also also known as the inferior temporal region (IT) in macaques. Experimental evidence strongly suggests that these areas process the object information and are responsible for their recognition. Cells within this pathway are busy with signals related to form, color and texture. The dorsal or "where" pathway extends from V1 and V2 to MT (middle temporal area) and MST (medial superior temporal area),

and continues to dorsal parietal areas. These areas are primarily concerned with the spatial information of images.

Other important pathways to mention are the magnocellular and parvocellular ones (also called M-pathway and P-pathway, respectively), each one of them carrying different kinds of information [22]. Early in the retina, large ganglion cells (magno, or M cells) project to the magnocellular portion of the lateral geniculate nucleus, that consequently sends neuronal projections to the superficial layer 4C in the primary visual cortex. This "system" is color blind, has high contrast sensitivity, fast temporal resolution, and low spatial resolution. Small ganglion cells in the retina (parvo, P cells) project to the parvocellular portion of the lateral geniculate nucleus, which projects to the deep layer 4C of the primary visual cortex. This system is color selective has low contrast sensitivity, slow temporal resolution and high spatial resolution.

Every stage within the ventral stream analyses the information with a progressive level of abstraction and complexity, in accordance with the growing receptive fields from V1 to TEO [23]. Cells in V1 serve as local spatio-temporal filters for orientation, spatial frequency and direction of motion. Receptive fields of simple cells in V1 respond to oriented stimuli, due to their linear on-and-off subregions. V2 cells with larger receptive fields may respond to virtual or illusory contours. From there the information follows to V4 where the majority of cells are responsive to contour features (angles and curves) [24].

The largest cell receptive fields found in the ventral stream are within the IT cortex in monkeys and LO (lateral occipital cortex) in humans [25, 26]. This area is more responsive to whole objects, primarily representing their shapes. Functional neuroimaging studies have evidenced important features of LO that allow these cells to detect and manipulate whole visual objects: (1) change in visual size does not affect it's activation, (2) high sensitivity to image scrambling, (3) convergence of visual cues, i.e. shapes can be defined by luminance, texture or motion, (4) invariance to changes in image position, (5) novel and memorized objects produce similar activations.

The existence of these pathways and "subsystems" within the visual system that communicate to each other clearly indicates that the brain is not only capable of managing local properties (i.e. orientation, texture analysis, edges) but also global properties of the visual scene (i.e. "where" it is in space and the object-to-context relationships). This is important to highlight, since any artificial vision system will necessarily have to be able to handle these two opposing but complementary portions of the visual information. That is to say, the artificial vision system should distinguish and split both types of information, the frequency related information and the object related information. A multi-resolution analysis would be an appropriate approach to do this and has been previously applied to some extent for the analysis of natural images and medical images as well [put references here].

2.2. Cognitive neuroscience of visual attention

Cognitive science has traditionally distinguished two varieties of attention. Overt attention refers to the rapid eye movements used to fixate the gaze on the attended object of the scene. It is not possible to direct gaze to one object without attending to it. Attention can also be deployed covertly, that is in a global manner, without looking directly to objects. Thus overt and covert attention represent the two ends of a spectrum of visual behavior.

Saccades are a type of eye movements that occur as the gaze shifts from one object to another in a scene. These movements are very fast and are interspersed by fixations of the eye that last between 200 and 300 ms. Fixations bring every new object of interest onto the fovea, a special zone of the retina located in the posterior pole of the eye. Due to some morphological and physiological characteristics, the highest spatial resolution (as high as 5.5μ or $1°$of the visual scene) of the incoming image is found in the fovea. The fovea is mainly constituted by cones, a type of retinal cells in charge of photopic vision, i.e. responsible for vision in bright light and color vision. The cone system has a greater acuity for resolving the details and boundaries of objects. The packing density of cones decreases towards the retinal periphery, where rods are the prevailing cells. Rods are more sensitive to light and subserve night vision. The rod system is not sensitive to color and has less resolving power than the fovea.

What is actually the focus of attention? What is brought onto the fovea due to overt attention mechanisms? Does covert attention select the objects in the scene in advance prior to be fixated by the eyes? These are intriguing questions and although the neural mechanisms of attention are still not completely understood, some explanatory models have been proposed [27,28]. One of the most influential models describes attention in the context of the nervous system's limited capacity of information processing and it's selectivity for information relevant to current behavior. The most important aspect of this model for the following discussion is that both, bottom-up, as well as top-down information serve as cues to resolve the competition between different objects to be attended.

There is consistent experimental evidence, that bottom-up saliency may be elicited through different kinds of cues, namely by single features (motion or color) [27, 29, 30], by objects [28,31–33] (conjunction of features), or by spatial location [34–36]. Importantly, this bottom-up driven saliency occurs in pre-attentive stages. In the case of single features, the fewer features an item has in common with the surrounding distracters, the more salient it is and easier to locate. In object-based selection, basic features and parts of objects can be integrated into whole structures, i.e. selection of only one visual attribute, enhances the representation of the other attributes of the same object automatically.

Top-down information biases the competition in favor of one of the multiple objects presented in the image and requires a previously stored representation of these objects as well as of the complete scene for further recognition in higher processing areas. As we mentioned in the previous section, object recognition processes occur within the ventral pathway. Neuropsychological and neuroimaging studies have evidenced the subprocessses of object recog-

nition [37] which can be split into four main stages: (1) extraction of basic features, (2) shape analysis, i.e. extraction of higher level information about the object, (3) matching to stored visual descriptions, (4) accessing of semantic/conceptual representations about the object.

All the mentioned processes depend on the reciprocal interaction between lower and higher order visual processing areas and do not necessarily occur in the order exposed. The extraction of basic features takes place within areas V1 and V2. Whole object shape analysis is accomplished in the IT cortex. Processes 3 and 4 depend on the connections between the temporal and prefrontal cortex. Additionally, a strong line of thinking suggests the existence of a category-specific, anatomically segregated modular organization of the IT cortex [38–40]. That is, some particular categories of objects (e.g. faces, houses, animals, tools, etc) are represented by the activation of distributed discrete areas, where semantic information about their features is stored. Important findings also reveal that IT cell's selectivity for specific object features can be modified through associative learning and that expertise with some categories enhances activity in the associated areas encoding their information [41, 42].

This brings us to the topic of "expertise", which is fundamental to understand within the scope of a highly trained medical specialty such as radiology. Expertise has also been studied extensively in cognitive science. Rosch et al. described three cognitive levels of an object or a scene description, i.e. superordinate, basic and subordinate levels [43]. These are three different levels of abstraction, with different semantic information which help to categorize between different objects and scenes. Experience determines the difference between basic and subordinate categories. With training, subordinate categories become new basic level categories. One of the most influential theories about the acquisition and evolution of expertise with specific categories of objects is the one proposed by Gauthier et al [41, 42]. Their proposal is supported by neuroimaging experimental evidence. This theory states that long and repetitive exposure to a specific category of objects (birds, cars, and eventually radiological images) enables neural automatic processes for immediate subordinate categorization and identification in the same manner as is done with faces. Face recognition is considered as a domain in which all humans are experts. Gauthier et al. demonstrated that the fusiform face area (FFA) which activates during face recognition tasks also serves as the neural substrate for object recognition in other expert fields of knowledge at an individual level.

The important idea to be emphasized concerning the evidence on expertise and face perception experimentation, is that face recognition differs from general object recognition because of "configural processing". Gauthier et al. [41, 42] argue that configural or holistic processing with faces is gained as people become experts with them permitting a subordinate identification of individual faces. That is, previous subordinate items (objects or scenes) are recognized as belonging to a basic category. This may occur in the same manner with novel objects as people become experts with them. As someone becomes expert in a specific domain, there will be a gradual shift from feature based to configural processing.

Holistic or configural processing refers to a representation and processing strategy, consisting in the integration of features into a gestalt or holistic representation. Integration of features implies that each face part cannot be processed independently from another, i.e. the recognition of one part is affected by the other face part(s). This is close to the idea of object based attention [28]. The evidence also suggests that the holistic information extracted from the image takes place at the perceptual encoding stage and not at the decisional stage [44]. Marr postulated in his influential model of vision that different spatial frequency bands enter the visual system through different channels (see section 1.1 on magnocellular and parvocellular pathways) and provided also different kinds of bottom-up information for perceptual and cognitive functions [45]. Low spatial frecuencies provide coarse visual information, i.e. large scale variations. High spatial frequencies represent fine visual information that relies on tighter luminance gradients. Interestingly, there is evidence that low spatial frequencies highly support configural processing in face perception [44].
Configural or holistic processing is intimately related to other cognitive theories about perception known for decades. These theories conceptualize holistic processing in terms such as schemata [46, 47], frames [48] or context frames [49], and are applied to the recognition of categories of objects, and natural scenes as well. In general, these theories rely on the idea that the identification of a particular object or scene initially depends on the construction of a schemata or context frame, which is activated by coarse global scene information. This kind of information arrives earlier to higher visual areas (V4, TE, TEO/IT), where the large receptive fields subserve covert attention, necessary at this initial moment. Thus, an initial approximation guess or the gist of the scene has to be reached. This initial gist has already semantic content and pertains to a specific basic level of categorization. Experiments have shown that visual scene semantics can be extracted in exposure times of around 100 ms, that is in the perceptual phase of encoding [46, 50].
After this initial phase, overt attention is required to confirm the guess. Overt attention is thus a phenomenon of feature analysis on the eye fixation points, taking advantage of the higher spatial resolution. The type of information required in this process arrives later and consists of the higher spatial frequencies. What is important in terms of expertise is that basic level categories of the observed object are detected at the first glance without requiring further processing. On the contrary, the identification of subordinate categories requires the incorporation of fine details, and thus further feature-based processing.

2.2.1. The role of working memory and the frontal lobe

According to Desimone et al. [28] two cues are necessary for visual attention, the bottom-up cue and a top-down cue, that modulates the activity in the IT cortex and biases the competition in favor of the most relevant stimulus representation. This theory states that the "focus of attention", requires a previously defined "attentional template", which is a sort of short description that represents any property of the relevant object. This template

may serve to monitor the object representation in higher visual areas of the cortex. It also has to be a rapid and temporary access to the semantic information and has to be activated immediately after the scene or object is presented or with any expectation of a specific object. As said before, the "attentional template" may also be interpreted as an analogy to the initial holistic representation that is activated with the coarse information provided by the low spatial frequencies that reach higher level cortical cells such as V4, IT, and TEO through the M visual pathway.
The most complex step of the visual attention process is the semantic conceptualization, which ultimately leads to the recognition and identification of the scene or object. Working memory is a complex form of very short-term memory which is constantly supporting other neural systems with important higher order information derived from the frontal cortex. Studies on visual working memory have demonstrated the existence of an interconnected system of different cortical areas that are able to carry out the temporary maintenance and manipulation of visual information in the absence of the visual stimulus. The involved areas are the IT cortex, the prefrontal cortex (PFC), and the medial temporal lobe.The three components of this system have been related to persistent activation during working memory maintenance of objects and spatial locations [51].
The most critical cortical area implicated in working memory is the prefrontal cortex. In general, the PFC has been related to cognitive control processes, by selecting or inhibiting relevant object representations, and by monitoring spatial and non spatial relations among items active in memory. Interestingly, the PFC region in the left hemisphere has also been implicated in semantic memory processes. Some investigators have suggested a "semantic working memory system" located in this region with strong influences over the temporal occipital cortex.
Important findings related to the working memory system is that the activation in IT cortex for a particular object, may be elicited by the expectation of it's appearance. That is to say, that there is an increased activity during directed attention in the absence of visual input and a larger increase after onset of the expected visual stimuli [52–54]. Activation within this area has also been linked to the maintenance of novel stimuli relative to familiar ones, i.e. they are involved in the retrieval of long term memory representations of objects, while associated cues are being presented. Ultimately, representations encoding object concepts in this latter region are selected, inhibited, manipulated and monitored via the PFC connections.

2.3. Visual perception of radiologists

Radiology is a medical discipline that depends on the interpretation of visual information provided by medical images that results in a reliable a diagnosis. Radiological interpretation is considered a domain of expertise [55]. Expertise is the ability to acquire and retrieve specific contextual knowledge that makes the difference between experts in different fields. Expertise is acquired by experience and in the case of radiology, expertise refers to reliably

solve diagnostic problems. Radiology trainees gain expertise by interpreting a vast amount of medical images during their training period or residency.

The radiological interpretative task has been studied under two approaches, a perceptual one and a cognitive one. The former approach analyzes eye-position data from the search for diagnostic information in an image. Answers in the cognitive approach are derived from the analysis of verbal protocols and sketches produced during the interpretation [55].

Research on the visual perception of radiologists has yield to important discoveries. It is now clear that experienced radiologists do not scan the whole image with a lot of fixations, but they rather bring up to the center of the gaze the most informative areas for diagnosis [56–58]. Although the highest resolving power resides within the fovea, most of the radiological image is not fixated and is left unexplored. Studies done on the interpretation of chest radiographs demonstrated that the detection of most abnormalities is reached between 10 and 20 sec or 30 to 60 fixations [59]. Experienced radiologists also tend to locate faster and more accurately the high informative areas compared to radiology residents. In fact, during training the visual scanning paths evolve from fixations concentrating on edges to more rapid accurate fixations on the abnormalities [57, 58]. The results from all the mentioned studies are consistent with the idea that speed and accuracy of abnormality detection determine expertise in radiology [60].

Resembling studies in cognitive psychology, experiments with radiologists have included protocols in which radiographs are presented shortly. These studies are also known as flash experiments and the experimental question is about how much the viewer (in this case a radiologist) can see in a glance. Previous cognitive studies in which a series of images is presented sequentially and the person is asked to only react to a specific target item (also called the rapid serial visual presentation paradigm) have demonstrated that observers can identify the category of a natural scene by extracting contextual information in about 100 ms [46, 50]. That is, semantic information is already available at this short time of exposure. Similarly, experiments with radiologists have demonstrated that with presentation times of 200 ms experienced radiologists identify 70% of the abnormalities. Interestingly, the recognized abnormalities are large, high-contrast targets, which significantly alter the normal anatomy of the image. The smaller and low-contrasted an abnormality is, the more time is needed to locate and recognize it [61, 62].

Other similar studies have investigated the perceptual schemas of radiologists by comparing their drawings of the displayed radiograph with the drawing of the same images done by laypersons [61]. These experiments demonstrated that the drawings of radiologists depicted actual objects identified within the image. The authors of the study concluded that the radiological schema consists of anatomical objects and the abnormalities are perceived as additional objects within this schema. These findings suggest that that the visual perception in radiologists is basically driven by whole objects, which is consistent wight the object-based attention theory previously exposed.

As mentioned before, the other approach for studying the radiologists' perception is the cog-

nitive approach. In these experiments, verbal protocols are recorded by having the radiologist read the radiograph while "thinking out loud" and followed by the dictated diagnostic report. Lesgold et al. studied expertise by comparing verbal protocols of radiologists with different levels of experience [63]. The analysis of verbal protocols led to a two stage diagnostic process. In the first stage a perceptual decision was taken on the basis of a set of perceptual features that characterized the image. The second stage consisted of a decision-making analysis of the perceptual features within a cognitive framework. Their important finding was that the first stage strongly depends on a schematic representation of the anatomy, i.e. a map of anatomic features. The more experienced the radiologist, the richer and more refined the anatomic schema. This schema maps the image features to the normal anatomy. The schema is also called up faster in experts. Features that do not match the schema provide signals for possible abnormalities.

Another important finding of the analysis of verbal protocols is that experienced radiologists tend to generalize the abnormal imaging findings from one particular case to idealized general patterns of pathology. Thus, each particular patient is projected to a "patient model" of anatomy, pathology and medical history. Evidence for this comes from a study that demonstrated that with the development of expertise in radiology, normal anatomical variants are selectively ignored, permitting a more refined normal and abnormal pattern construction and detection of real perturbations [64]. This finding supports the idea of the existence of different patterns for each group of pathology in addition to the normal pattern.

Radiologists also make use of several and different cues to detect the abnormalities in an image. Cues are hints about the image that are part of the pictorial content. There are external and internal cues. External cues are arrows or circles that indicate probably important diagnostic locations within the image, such as the ones used in Computer Aided Diagnosis (CAD) systems. Internal cues are those image findings that suggest the existence of other findings. It has been shown that the radiologist's search pattern is highly dependent of what is seen during the scanning of the image. That is, the first detected and recognized abnormality will serve as a suggestion for where and what to look for later [65].

To summarize, the expert performance in radiology is gained during the residency training and is intimately related to the building and further rapid activation of a global schema of the different types of radiological images (CT, MRI, X-ray, etc). The ultimate goal of this global schema is to determine wether an image is normal or abnormal. This schema represents the normal anatomy, and at the first glance oddities or perturbations to the normality are localized. The most conspicuous perturbations are then focused and analyzed in detail, and will serve as a guide for the following search strategy. This sequence of events led to the formulation of the radiologist's perception called the *global-focal model* [61,65,66]. According to this model the image is analyzed with two perceptual and cognitive strategies, one global strategy and one focal strategy. The global strategy uses information from all the retina including it's peripheral portions and generates a general glimpse of the image and establishes the major spatial relationships between objects. As the experimental evidence

shows, radiologists gain and tune the global analysis during their training. Global analysis alternates with focal analysis, which depends on more resolving power for feature-based analysis of the fixated object. It is a more detailed processing of the object projected on the fovea.

2.4. Previously developed models of visual attention

As a result of the neurophysiological and cognitive evidence on visual attention and the mammal visual system in general, the computational models that have been proposed and developed so far have either a feature-based or an object-based approach. As part of these models, bottom-up or top-down computation strategies have also been included in either a pure form or as a combination of both.
The concepts of top-down and bottom-up information flow are based on previously exposed neurophysiological theories and experimental evidence. Bottom-up models have been more extensively developed since there is more evidence on its neurophysiological correlates in the brain and the easiness of the implementation. The primary assumption of bottom-up models is that visual information is endowed with inherent low level characteristics that are processed in a pre-attentive manner. So called "pop-out" effects and perceptual grouping phenomena may occur within this pre-attentive period. These self organizing properties are supposed to be enough to guide the focus of attention. Some pure bottom-up models also consider that these low level characteristics control the selection, movement and inhibition of one attended region to another, resembling a serial "scanpath" of attention. On the other hand, top-down approaches take into account that internal representations of the outside world guide the focus of attention. Thus, top-down phenomena are more related to volition (voluntary actions), knowledge of a specific task, the semantic category of the scene (the gist of the scene) and prior knowledge of the target. . Many models combine both types of information control. Lately the efforts have been also directed towards the coupling of attention and visual recognition algorithms.
One of the most influential computational models of bottom-up attention was proposed by Koch and Ullman [67]. This model is based on a psychophysical theory of attention known as "feature integration" [27], that was derived from visual search experiments. This theory postulates that visual objects are characterized by very basic dimensions such as orientation, size, color, closure, intensity, flicker and direction of motion. Objects automatically "pop-out" from the image if they differ from the surrounding objects (that act as distracters) in only one single features. If an object is distinguished from the distractors by a combination of single features (a conjunction of features), then the "pop-out" effect takes longer because a direct attention deployment is needed, permitting the target's identification and location. If there are more than two objects that are distinguished from the rest by a combination of features, then attention must be directed serially to one object at a time. This proposal led to the idea of considering attention as a "spotlight" that serially illuminates each object of

a scene.

Based on this psychophysical theory Itti and Koch developed a computational model [1]. The key component of their model is the saliency map, which is the final pathway after single feature maps (e.g. for color, contrast and intensity) have been initially computed in parallel and have competed between each other to become salient. At the end, an external winner-take-all algorithm, which can be interpreted as the top-down modulation, selects in a serial manner the more salient locations within the saliency map.

Figure 2.1 shows the complete diagram of the model by Itti et al. and a detailed description of the computation of the model has been described in [1]. Initially, a color image is processed with a linear lowpass filter consisting of Gaussian pyramids that create nine spatial scales of the same image. These low frequency maps of the image serve as input for the calculation of the saliency map. The computation is based on center-surround operations that mimic processes that take place in the retina, lateral geniculate nuclei and V1 cells. These cells function as linear filters for the three basic image features (intensity, color and orientation). The center-surround architecture is implemented as the difference between the fine and coarse scales. Seven center-surround feature types are used: on/off image intensity contrast, red/green and blue/yellow double opponent channels, and four local orientation contrast channels. For each of these contrast types six different feature maps are computed at different pairs of center and surround spatial scales to finally get 42 different feature maps. Afterwards, the feature maps are normalized by a factor that enhances the locations with strong peaks of activity and suppresses those with multiple comparable peak responses. This procedure is supposed to resemble the lateral inhibition mechanisms of cortical networks. Subsequently, all the feature maps for each characteristic are added across the different scales, by initially reducing each map to the corresponding scale and adding them point-by-point. This addition yields to the saliency map for each feature. Finally, the three saliency maps (intensity, color and orientation) are normalized and summed together into a generalized saliency map.

Attempts to integrate top-down information into Itti's saliency-based model include the one by Rapantzikos et al. [68]. Their approach integrates prior knowledge and motion processing to the original model scheme in order to draw attention to faces in video coding. They make use of a traditional color based skin detector and a multiresolution gradient-based approach to estimate optical flow and run them as additional channels next to the basic features of color, intensity and orientation. This yields to a total of five conspicuity maps, which are then combined into the final saliency map that locates faces as regions with higher activity in the image.

Itti's group enhanced their own model by adding top-down cues to locate the target of an image [69]. This enhanced model gives a task in the form of a set of keywords. These keywords are then analyzed by an ontology that outputs task-related entities and their relationships. The model then looks for the most relevant task-related entity in the visual scene which is then considered a target. The low-level visual features processed through Itti's model are biased according to the known features of the target in order to make the target

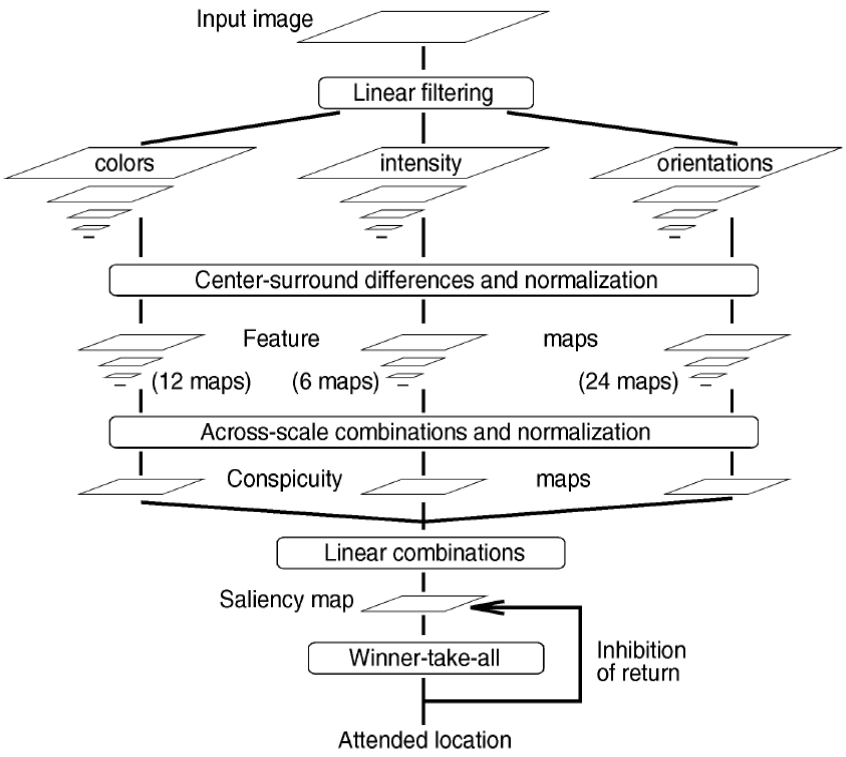

Figure 2.1.: Itti's Model of Visual Attention. Taken from [1]

more salient. This closes the top-down to bottom-up loop. Additionally, the authors added a recognition module which then links the low-level visual information to the task-relevancy features and ultimately generates a task-relevance map that is constantly feeding a memory module.

There is enough evidence of object-driven attention in monkeys as well as in humans and some cognitive theories have been proposed that argue that the selection of objects is a process of parallel competitions of their combined neural representations in the scene. Thus, models following this neurophysiological/cognitive theory focus on the importance of whole objects as the directing elements of attention. Contrary to the serial processing exposed by the feature integration theory, this model proposes a parallel inflow of low level information to primary and higher visual areas. These low-level descriptors are grouped together, thus constituting themselves as integral objects. Objects compete for the limited processing capacity in the hierarchically higher visual areas, that are densely connected with prefrontal regions. In these models, competition is biased by bottom-up cues, as well as by top-down ones, such as behavioral and "attentional templates" stored in working memory.

Stark and colleagues have proposed a purely top-down attentional gaze shift model [70]. The model is based on the "scanpath theory" [71], in which the sequence of eye fixations is under the control of an internal cognitive model. Their model selects the most relevant sites of a scene based on prior knowledge of similar scenes. This is achieved by two modules, a

learning one and a cognitive one. In the learning module, the image is initially segmented into regions. A probability is given to each region of being a hypothetical object from a specific scene category. The recognition module shares the initial processing with the learning one up to the description of the segmented descriptors. By using a bayesian approach they estimated the probability of a scene category to contain a specific object.

Deco and colleagues have proposed a combined recognition and attention model [72]. Their model is inspired by the biased competition theory, stresses mechanisms for space-based and object-based attention, involves interactions between the dorsal and ventral visual streams, and includes top-down and bottom-up modulation strategies. They use neural networks models, each of them representing a cortical visual area. Within each module a competitive network is implemented by local lateral inhibitory connections. Modules are hierarchically interconnected by feed-forward and feedback connections. The input image enters the V1 module, which has hyper-columns of Gabor-modeled neurons to extract location, orientation, symmetry, and spatial frequency. In general, the visual search, as well as the object attention and object recognition are accomplished by introducing top-down biasing cues from the higher processing modules that ultimately have back-projections to the lower visual areas (V4-V2-V1).

Another approach, combining attention and recognition was proposed by Rybak et al. [73]. This model is also built in the context of the "scanpath theory" previously exposed [71]. Thus, it depends on an internal model of objects that directs eye fixations based on motor and sensory representations stored in memory. A low-level subsystem receives the input image and decreases its resolution from the center to the periphery in each fixation point. Afterwards edges are extracted in a twofold manner, i.e. a basic set of edges at the center and 'context edges' at the periphery. A mid-level processing module transforms these primary features in second-order invariant features. This is accomplished with the relative orientations and relative angular locations of the context edges with respect to the basic edges. The high-level subsystem functions in three different modes: memory, search and recognition. Initially, each image portion at each fixation point is stored in a sensory memory module and each following fixation position is stored in a motor memory module. In the search mode each new fixation point is compared to the stored images. If a match occurs, the recognizing mode executes the consecutive fixation movements according to the previously memorized patterns.

Some object-based approaches have been implemented with basic visual feature extraction algorithms and saliency maps. For example, Sun et al. [74] proposed a model integrating competitive interactions of objects and locations. As with feature-based models, this one initially extracts basic features (color, intensity and orientation), which then are grouped. This grouping is based on Gestalt theory principles and on heuristic knowledge, that is, proximity, closure, continuity, common fate, familiarity, and shared properties. Finally, spatial and grouping saliency maps are constructed and a top-down biasing module selects the fixation points.

Han et al. [75] extract objects from a visual scene after Itti's saliency-map is computed. The authors also compute a texture-based edge detection algorithm. By maximizing the most salient areas in the saliency-map and by minimizing the local edge values, the model is able to selectively seed the scene. The location of the seed is considered as an attention object which then grown by means of a markov random field model. This last step is implemented serially in terms of the decreasing order of saliency value.
Orabona et al. [2] created a very sophisticated visual attention model that was implemented on a robot, with bottom-up and top-down cues as well as attention and recognition modules. This model initially extracts basic features from a color image by extracting four color channels (red, green, blue and yellow) and by computing center-surround receptive fields on each pixel for red-green, green-red and blue-yellow opponencies. Edges are then calculated for each of the three color-contrasts by using a Sobel filter. The edge maps are then combined in order to generate a generalized edge map which afterwards undergoes a watershed algorithm in order to extract so called "proto-objects". The watershed algorithm fills out the spaces in between the edges ultimately leading to segmented blobs that are tagged based on the average color inside it. Bottom-up saliency is computed as the euclidean distance in the color opponent space between each blob and its surrounding. Top-down influence is calculated based on a specified task which biases the bottom-up saliency in favor of a stored representation of the target to be searched. The general structure of the model can be seen figure2.2.
In summary, a perfect model of visual attention should be able to resemble most of the visual system's functionality, but this is realistically speaking beyond the capabilities of the all the models presented here. When choosing an approach, one should look for the specific features of the of the target images (task-related choice) and also learn from particular fields of expertise in order to "shape" one's model according to it. Something that is be very important and should be part of a model trying to emulate the radiologic diagnostic process is the bidirectional flow of information. The communication between modules processing different types of visual information (top-down and bottom-up processing) in some way ensures that only the most relevant information will be captured by reinforcing and inhibiting it. It is also important to split the visual information, for example in higher and lower spatial frequencies (magno and parvocellular pathways), and have each module process it separately. This multi resolution approach is very close to what has been seen on actual neurons of the visual system. Finally, higher complex semantic information is a very sophisticated way to actually categorize and ultimately identify the extracted objects in the scene (MS lesions in this case). These "semantic modules" emulate the the functions exerted by the frontal lobe in the way that working memory does.

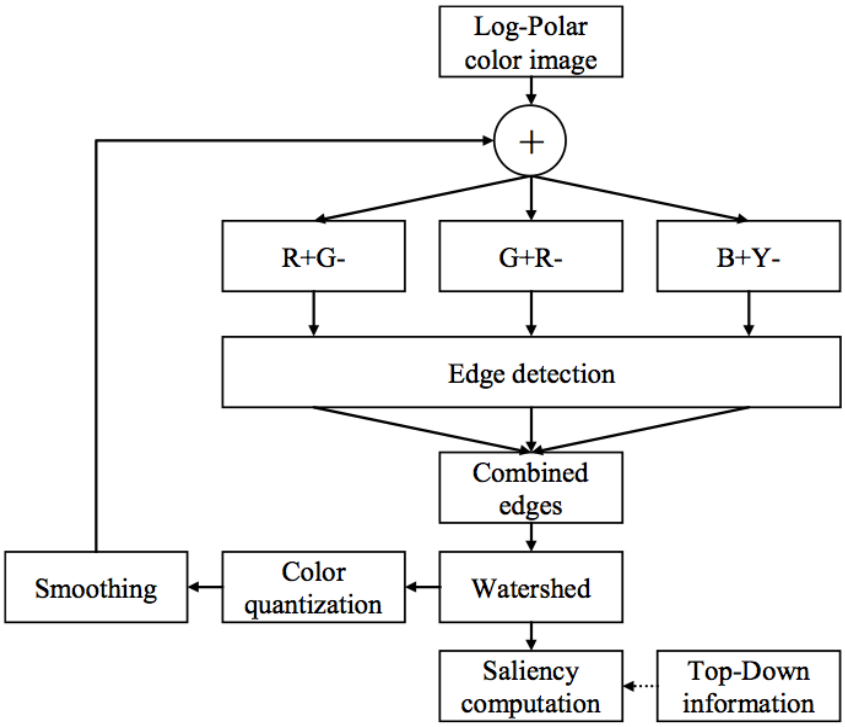

Figure 2.2.: Orabona's object-based model of visual attention. Taken from [2]

3. A fully bioinspired model of visual attention

3.1. Texture border processing model

The model proposed here is designed to characterize the relevant diagnostic regions in pathologic magnetic resonance images (MRI) of the brain, by trying to resemble the neural mechanisms that happen during the radiological interpretation process. The model attempts to include bottom-up and top-down processing streams as described in Desimone and Duncan's theory of attention [28]. In addition to this, it also focuses on the analysis of texture in the image as it is proposed in the visual model proposed by Thielsher et al. [3,4]. It is important to mention that the initial part of the radiological diagnostic process (searching and detecting the pathologic regions in the image) is highly dependent on the contrast produced by either luminance as well as by texture that distinguishes boundaries between normal appearing tissue and pathology. Luminance contrast has been more widely studied in vision research than texture contrast. Also, to the extent of our knowledge, there are no studies that investigate the radiological perception based on texture when interpreting MRI images. The reasons we explored the usage of the proposed model is based on the fact that boundary detection is also possible via texture analysis and that the texture border detection is mainly achieved by cells in higher model areas. The latter property serves our goals because it will allow us to explore the interaction between hierarchically lower and higher areas in the visual ventral stream (bottom-up and top-down processing). The original model proposed by Thielsher et al. will be first described below, followed by the model proposed in this work, which is a variation of it.

The visual system utilizes several different basic image features to group the incoming visual stimulus into distinct objects. These features are: luminance, color, texture, motion and stereoscopic depth. Higher order visual areas such as V4 and TEO require the image to be segmented into objects which can be detected by following the discontinuities inherently provided by these basic features. In other words, the detection of borders is crucial for the segmentation and identification of objects. In the case of radiological images (X-rays), luminance and texture contrasts have been used to detect, characterize and segment relevant diagnostic regions [76]. Despite the fact that the physiological basis of luminance contrasts guiding the detection of borders is better understood than that of texture, here we focus on texture border detection. This strategy also allows us to test the importance of texture for

the radiological interpretation process and also evaluate it's contribution to the "global-focal" model (See section 2.3 for more information).
The model is built by a set of hierarchically organized model-cells (i.e. filters) that represent specific groups of cells with similar activation patterns. These cellular groups resemble the different stages or steps of the the visual system's information flow. These model-cells are: the lateral geniculate nucleus (LGN), V1 simple cells, V1 complex cells, area V2 and area V4. The three latter ones, i.e. V1 complex cells, area V2 and area V4, are bi-directionally connected to each other resembling the feed-forward (bottom-up) and the recurrent (top-down) processes and connections in the brain (See figure 3.1). Unless otherwise stated, these filters will be referred as "cells" for now on in this dissertation. V1 complex cells , V2 cells and V4 cells constitute the three higher stages of the model and each of them undergoes three successive "activation stages": 1) pooling of bottom-up activity, 2) activity modulation via feedback interaction (top-down modulation), 3) intra-areal center-surround competition. (See figure 3.2).

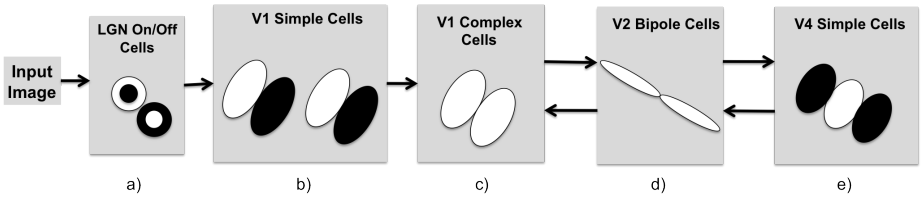

Figure 3.1.: Thielsher's Model of Texture Boundary Detection. Taken and modified from [3]

The first activation level of V1-complex, V2 and V4 cells is the bottom-up activation, which is modeled as a linear equation that implements a spatial convolution and pools the input into the cell's receptive field. During the second level cell activation the initial bottom-up input is modulated via the feedback signal originated in higher model areas (top-down modulation). In the third and last level of activation, the top-down modulated activity undergoes center-surround (ON-center/OFF-surround) competition between cells of their spatial and orientational neighborhood (See figure 3.2).

3.1.1. Top-down modulation and Center-surround competition

Bottom-up input pooling, top-down modulation and center-surround competition are the three steps of activation that characterize the behavior of V1 complex cells, V2 cells and V4 cells. The dynamics of the second (top-down modulation) and third (center-surround) steps of activation will be explained first, because they are common to the three cells, whereas the dynamics of the cells' first step of activation (bottom-up input) is unique to each cell and will be explained separately in the next section (section 3.1.2).
The dynamics of the top-down modulated activation (Figure 3.2, section 2) is modeled as a differential equation:

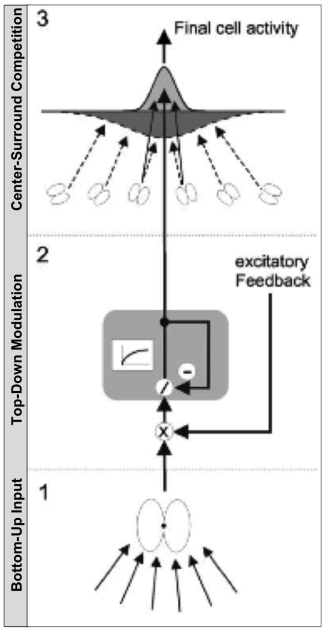

Figure 3.2.: Activation steps for each filter (model cell) in the attention model. Taken from [3]

$$\frac{\partial}{\partial t} I_{i\theta}^{(1)} = -\alpha_1 I_{i\theta}^{(1)} + (\beta_1 - \gamma_1 I_{i\theta}^{(1)}) c_{i\theta}[1 + C h_{i\theta}] \tag{3.1.1}$$

where $c_{i\theta}$ is the bottom-up input that comes from the first activation process (see section 3.1.2) and is sensitive to location i and orientation θ, whereas $h_{i\theta}$ is the top-down feedback of the cell's activity located in a higher stage of processing. The equilibrated response of equation 3.1.1 when $\frac{\partial}{\partial t} I_{i\theta}^{(1)} = 0$ is:

$$I_{i\theta}^{(1)} = \frac{\beta_1 c_{i\theta}[1 + C h_{i\theta}]}{\alpha_1 + \gamma_1 c_{i\theta}[1 + C h_{i\theta}]} \tag{3.1.2}$$

where β_1 and C control the strength of the excitatory feedback (top-down modulation). α_1 determines the rate of activity decay. Afterwards, $I^{(1)}$ becomes the input to the third computational stage, which exerts a scheme of shunting center-surround competition (ON-center/OFF-surround) (Figure 3.2, section 3). This stage is expressed by another differential equation:

$$\frac{\partial}{\partial t} I_{i\theta}^{(2)} = -\alpha_2 I_{i\theta}^{(2)} + \beta_2 \{I^{(1)} * \Psi^+ * \Lambda^+\}_{i\theta} - (\delta_2 + \zeta_2 I_{i\theta}^{(2)}) \{I^{(1)} * \Psi^- * \Lambda^-\}_{i\theta} \tag{3.1.3}$$

In this equation, Ψ^+,Ψ^-,Λ^+ and Λ^- denote gaussian weighting functions for the excitatory and inhibitory activity and "$*$" denotes the convolution operator. Interestingly, this equation has the same form as a "mexican hat" for spatial and orientational information. The

mexican hat is a non-orthogonal, symmetric, crude wavelet. This means that it is defined by a mathematical expression that draws a continuos and infinite waveform and by using equidistant discrete points along this curve we can create its corresponding filter. The mathematical expression that defines the "mexican hat" wavelet filter is the second derivative of the Gaussian probability density function. It can also be defined as an operator that applies a two dimensional Laplacian to the image. This operator is the core of the Marr-Hildreth edge detector that was very commonly used for edge detection in image processing [77].
The equilibrium state of equation 3.1.3, when $\frac{\partial}{\partial t} I_{i\theta}^{(2)} = 0$ derives into:

$$I_{i\theta}^{(2)} = \frac{\beta_2\{I^{(1)} * \Psi^+ * \Lambda^+\}_{i\theta} - \delta_2\{I^{(1)} * \Psi^- * \Lambda^-\}_{i\theta}}{\alpha_2 + \zeta_2\{I^{(1)} * \Psi^- * \Lambda^-\}_{i\theta}} \tag{3.1.4}$$

This last equation shows how the signal $\beta_2\{I^{(1)} * \Psi^+ * \Lambda^+\}$ is interacting with inhibitory substractive activity in the neighborhood represented by $\delta_2\{I^{(1)} * \Psi^- * \Lambda^-\}$ and by shunting divisive inhibitory activity represented by $\zeta_2\{I^{(1)} * \Psi^- * \Lambda^-\}$. At a glance, the model initially extracts salient texture arrangements that are then modulated with top-down feedback information and ultimately the activity is followed by center-surround competition where local irrelevant information is suppressed within the neighborhood, which is what is expressed in equation 3.1.4.

3.1.2. Bottom-up input

The second and third activation levels of V1 complex cells, V2 and V4 cells were described in the section above. We will now describe the initial bottom-up activation of all the cells that are part of the model (LGN, V1 simple cells, V1 complex cells, V2 cells, V4 cells). The only pure feedforward model cells are the first two cells in the model, LGN cells and V1 simple cells (See Figure 3.1 *a*) and *b*)), which means that they do not undergo the second and third stage of activation explained in section 3.1.1. These two first cells greatly mimic the parvocellular visual pathway and in this model they are the only ones that filter the input's luminance distribution. The LGN cell uses circular center-surround receptive fields to detect local luminance transitions:

$$x = I * (\Lambda_{Center} - \Lambda_{Surround}) \tag{3.1.5}$$

$$x^{on} = [x]^+$$
$$x^{off} = [-x]^+$$

Here, I is the input image, which is convolved by the difference of isotropic 2D gaussian kernels represented by Λ_{Center} and $\Lambda_{Surround}$. $[x]^+$ and $[x]^-$ represent half-wave rectifications. Beware that this center-surround competition is different than the one described before in equation 3.1.4. The input to the V1 simple cells are the bottom-up ON and OFF activations

of LGN (represented by x^{on} and x^{off}). V1 simple cells exist for two polarities (dark-light: dl; light-dark: ld) and eight orientations. That is, V1 simple cells are the first ones to process not only luminance information but also texture information. As it is shown in part b) of figure 3.1, these cells filter the image with elongated ovoid ON (light) and OFF (dark) subfields:

$$p_\theta^{on/off-left} = x^{on/off} * \Lambda_{\sigma_x,\sigma_y,0,-\tau_y/2,\theta}$$

$$p_\theta^{on/off-right} = x^{on/off} * \Lambda_{\sigma_x,\sigma_y,0,\tau_y/2,\theta}$$

Here, p_θ denotes the subfields of V1 simple cells, $\Lambda_{\sigma_x,\sigma_y,0,\pm\frac{\tau_y}{2},\theta}$ are 2D anisotropic gaussian weighting functions, in which the standard deviations σ_x and σ_y define the size and shape of the subfield. θ determines the orientation of the subfield. The subfields are initially shifted perpendicular to their axis by $\pm\frac{\tau_y}{2}$ and rotated by θ, that has eight orientations ($n_{orient} = 8$), so that $\theta = 0, \pi/n_{orient}, ..., (n_{orient} - 1)\pi/n_{orient}$. Thus, the analysis is split into eight maps, one for each orientation θ and it will be conducted in this manner throughout all the upcoming steps in the higher order cells (V2 and V4). The subfields $p_\theta^{on/off-left}$ and $p_\theta^{on/off-right}$ feed the activation of V1 simple cells S in a specific spatial location i and for each orientation θ. For example the activation that is selective for light-dark polarity is:

$$S_{i\theta}^{ld} = \frac{A_s(p_{i\theta}^{on-left} + p_{i\theta}^{off-right}) + 2B_s p_{i\theta}^{on-left} p_{i\theta}^{off-right}}{A_s D_s + E_s(p_{i\theta}^{on-left} + p_{i\theta}^{off-right})} \tag{3.1.6}$$

and the activation that is selective for dark-light polarity is:

$$S_{i\theta}^{dl} = \frac{A_s(p_{i\theta}^{off-left} + p_{i\theta}^{on-right}) + 2B_s p_{i\theta}^{off-left} p_{i\theta}^{on-right}}{A_s D_s + E_s(p_{i\theta}^{off-left} + p_{i\theta}^{on-right})} \tag{3.1.7}$$

In these last two equations, A_s, B_s, D_s, E_s are factors that control the impacts of the additive (left side of the numerator), multiplicative (right side of the numerator), and divisive components of the subfield responses.

After the initial activation of LGN and V1 simple cells, the model is characterized by the bidirectional traffic of information (feed-forward and feed-back) across the "higher order" cells (V1 complex cells, V2 bipole cells, and V4 cells) (See Figure 3.1 *c*)). V1 complex cells pool the activity of two simple cells of opposite polarity ($S_{i\theta}^{ld}$ and $S_{i\theta}^{dl}$) at each position i by calculating a half-wave rectification from the differences between the two activations of opposite polarities ($S_{i\theta}^{ld}$, $S_{i\theta}^{dl}$):

$$c_{i\theta}^{V1} = A_c([S_{i\theta}^{ld} - S_{i\theta}^{dl}]^+ + [S_{i\theta}^{dl} - S_{i\theta}^{ld}]^+) \tag{3.1.8}$$

V1 complex cells' activity is already sensitive to the orientation contrast, which means that it highlights areas where there are notable orientation changes from one receptive field to the other. The first activation of V1 complex cells ($c_{i\theta}^{V1}$) then undergoes the second activation ($I_{V1}^{(1)}$) with top-down modulation (see equation 3.1.2) and then the third activation ($I_{V1}^{(2)}$) of within-area center-surround competition as expressed in equation 3.1.4.
Thereafter, V2 cells take as input the last activation ($I_{V1}^{(2)}$) of V1 complex cells (See Figure 3.1 part *d*)). $I_{V1}^{(2)}$ is a 3D matrix, that encodes a 2D spatial matrix and a 1D orientational matrix ($\theta = 8$) and it is pooled into left and right subfields of activations represented by f:

$$f^{left} = I_{V1}^{(2)} * \Psi_f * K^{left} \tag{3.1.9}$$

$$f^{right} = I_{V1}^{(2)} * \Psi_f * K^{right} \tag{3.1.10}$$

$$f^{left} = c_{i\theta}^{V1} * \Psi_f * K^{left} \tag{3.1.11}$$

$$f^{right} = c_{i\theta}^{V1} * \Psi_f * K^{right} \tag{3.1.12}$$

Here, the orientation domain is blurred via the convolution with the isotropic gaussian kernel Ψ_f and in the 2D spatial domain with the anisotropic gaussian weighting function $K_{i\theta}$. These are elongated subfields which are shifted parallel to the main axis of the cell and are cut off in the central part of the cell by means of a sigmoid function. $K^{left/right}$ are represented by:

$$K_{i\theta}^{left} = \Lambda_{\sigma_{kx},\sigma_{ky},\tau_{kx},0,\theta}(\vec{X}_i) \times \frac{1}{1 + exp(-A_k \vec{X}_i \binom{cos\theta}{sen\theta} - B_k)}$$

$$K_{i\theta}^{right} = \Lambda_{\sigma_{kx},\sigma_{ky},-\tau_{kx},0,\theta}(\vec{X}_i) \times \frac{1}{1 + exp(+A_k \vec{X}_i \binom{cos\theta}{sen\theta} + B_k)}$$

The activities f of the V2 cells are combined in order to obtain the initial activation $c_{i\theta}$ of V2:

$$c_{i\theta}^{V2} = \frac{A_t(f_{i\theta}^{left} + f_{i\theta}^{right}) + 2B_t f_{i\theta}^{left} f_{i\theta}^{right}}{A_t D_t + E_t(f_{i\theta}^{left} + f_{i\theta}^{right})} \tag{3.1.13}$$

This equation is very similar to equations 3.1.6 and 3.1.7, in which the activities of opposite subfields (left and right in this case) are weighted by their additive and multiplicative interactions. $c_{i\theta}^{V2}$ goes then throughout its second and third activation levels ($I_{V2}^{(1)}$ and $I_{V2}^{(2)}$). The final step of processing is the model cell V4, which is hierarchically the highest cell of the model (See Figure 3.1 part *e*)). This cell measures the differences between the final activations of V2 cells ($I_{V2}^{(2)}$) that are pooled into an excitatory center field (q_{φ}^{center}) and

left and right inhibitory subfields (q_{φ}^{left} and q_{φ}^{right}). The center field (q^{center}) is defined as a gaussian kernel:

$$q_{\varphi}^{center} = I_{V2}^{(2)} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,0,\varphi} \tag{3.1.14}$$

and the left and right subfields are also defined as anisotropic gaussian kernels:

$$q_{\varphi}^{left} = I_{V2}^{(2)} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,-\tau_{qy},\varphi} \tag{3.1.15}$$

$$q_{\varphi}^{right} = I_{V2}^{(2)} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,\tau_{qy},\varphi} \tag{3.1.16}$$

"$*$" denotes the convolution operator. The rotation angle of the subfields is φ and the inhibitory subfields q_{φ}^{left} and q_{φ}^{right} are shifted perpendicularly to their main axis by $+/-\tau_{qy}$. V1 and V2 cells processed the information in eight different orientations ($\theta = 8$), and V4 assigns to each orientation θ eight orientations ($\varphi = 8$). This creates a matrix of $8 \times 8 = 64$ V4 cells. The initial activation of V4 is then calculated as the sum of the two half-wave rectified differences:

$$c_{i\theta}^{V4} = [q_{i\theta\varphi}^{center} - Cq_{i\theta\varphi}^{left}]^{+} + [q_{i\theta\varphi}^{center} - Cq_{i\theta\varphi}^{right}]^{+} \tag{3.1.17}$$

As before, θ denotes the orientation of the V1 and V2 cells, but φ denotes the orientations of the V4 cells. What equation 3.1.17 is saying is that for each V1/V2 cell orientations θ there is a group of V4 cells interacting between each other in a spatial and orientational neighborhood (eight total and denoted by φ). Each V4 orientation φ evaluates the orientations in the θ domain, suppressing those areas where there are homogeneous distributions of texture (texture elements with the same orientation pattern) and highlighting the texture borders (areas with huge gradients of orientation activity). V4 is the cell at the highest level of processing and it is not modulated by top-down activity of any other higher order cell. Thus, $c_{i\theta}^{V4}$ immediately undergoes the inter-areal center-surround competition (third level of activation as expressed in equation 3.1.4) right after the first bottom-up activation. Once all responses of V4 ($I_{V4}^{(2)}$) for each V1 orientation θ are calculated, they are summed up together (see next equation) and then sent back to modulate the activity of V2's second activation level:

$$I_{V4i\theta}^{(2)} = \sum_{k=1}^{n_{orient}} I_{V4i\theta(k-1)\times\pi/n_{orient}}^{(2)}$$

Figure 3.3 shows the performance of Thielscher's model on an artificial test image.

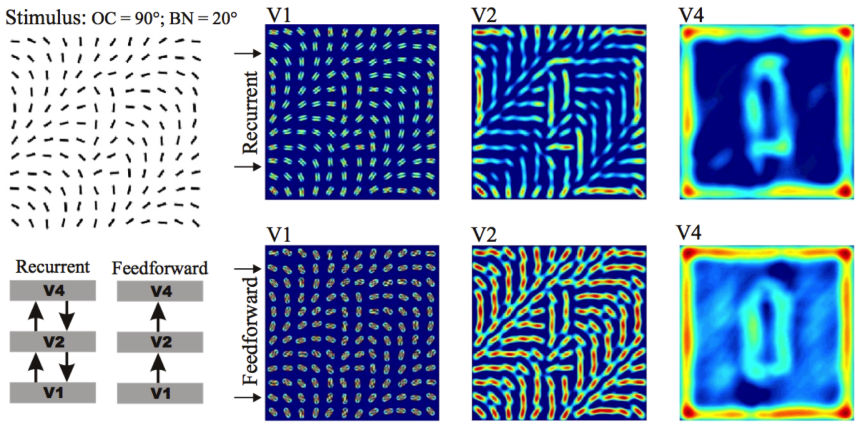

Figure 3.3.: This figure shows and example of the results of the equilibrated responses of the three higher order model cells as presented in the original model proposed by Thielsher et. al. 2003. The image shows a central pop-out bar surrounded by background noise and the performance of two different versions of the model, a recurrent version (with top-down modulating activity) and a purely feed-forward model. It is clearly shown that the recurrent model is able to better highlight the contours of the central bar as well as to suppress the background noise. Figure taken from [4].

3.2. The modified texture processing model

The model implemented here is a variation of the original model proposed by Thielscher et al. that was described above. Despite this, we did not change any of the core equations of the model. Most of the coefficients, multiplicative and divisive factors that regulate the activity within each cell's activation step were left the same as specified in the paper of Thielscher et al. of 2003 [4]. All of these parameters were previously proven to satisfy the stability of the system. We modified two coefficients - β and C - of equation 3.1.2. In this equation, these two parameters control the strength of the top-down modulation of $h_{i\theta}$ (activity of the higher level cell) over $c_{i\theta}$ (activity of the lower level cell). In our experiments we found better results when doubling the value of these two parameters. This was particularly true for the top-down modulation exerted by the V2 cell to the V1 complex cell (See the interaction between these two cells in figure 3.1). In addition to this, we also tried different combinations of gaussian kernel sizes for the bottom-up activation of each model cell. Table 3.1 shows the different gaussian kernel widths used throughout the model for each cell along with all the other parameters specified in equations 3.1.1 through 3.1.17. In general, we chose the combination of gaussian kernel array sizes that provided with the best results in terms of characterizing the multiple sclerosis lesion areas in the MRIs of the brain. The scheme that worked best consisted of growing gaussian kernel sizes from LGN (smallest size) to the higher order areas (V4 cells had the biggest size). This may be related to the fact that the receptive

fields increase in size along the visual system when moving from the lateral geniculate nuclei and striate cortex (V1) to higher order neuronal assemblies (V4).

	Parameters of $I^{(1)}$				Parameters of $I^{(2)}$							
	α_1	β_1	γ_1	C	α_2	β_2	δ_2	ζ_2	σ of			
									Ψ^+	Λ^+	Ψ^-	Λ^-
V1	12.0	1.46	3.7	50	1.0	2.8	3.5	5.0	0.2	1.0	2.0	3.0
V2	12.0	0.73	4.2	25	1.0	2.9	3.1	50.0	0.2	2.0	2.0	6.0
V4	-	-	-	-	1.0	10.6	9.9	1000.0	0.2	8.0	2.0	24.0

Table 3.1.: This table shows the parameters used for the second and third levels of activation ($I^{(1)}$ and $I^{(2)}$ respectively) and the respective gaussian kernel widths for each model cell.

A brief view of the model implemented here is shown in the figure below (Figure 3.4). Here, the input image is initially processed by LGN ad V1 simple cells by using the luminance distribution of the image. V1 simple cells process this information along 8 different orientations for either light-dark or dark-light polarities. The output of V1 simple cells' bottom-up activity is then sent to V1 complex cells (this sequence of events can be seen in figure 3.4 parts *a*), *b*) and *c*) and with more detail in figure 3.5 part *D*.). After this point our model starts differing from the original one proposed by Thielscher et al. We tested different forms of feed-forward and feed-back interactions between the higher order cells (Figure 3.4 parts *c*), *d*) and *e*)). What is shown in the bottom row of figure 3.4 is the final arrangement of interactions that resulted in the best results.
V1 complex cells pass over the output of their bottom-up activation to V2 and V4 model cells. Then, V2 and V4 start off with their own bottom-up activations. V2 cells get activated as in equation 3.1.13, but in this case the input to the the left and right hemi-fields is the first activation of V1 complex cells $I^{(1)}_{V1}$:

$$f^{left} = I^{(1)}_{V1} * \Psi_f * K^{left} \tag{3.2.1}$$

$$f^{right} = I^{(1)}_{V1} * \Psi_f * K^{right} \tag{3.2.2}$$

In the case of V4, the first activation is exactly as stated in equation 3.1.17 but the excitatory center field and the inhibitory left and right fields are also fed with $I^{(1)}_{V1}$:

$$q^{center}_{\varphi} = I^{(1)}_{V1} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,0,\varphi} \tag{3.2.3}$$

$$q^{left}_{\varphi} = I^{(1)}_{V1} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,-\tau_{qy},\varphi} \tag{3.2.4}$$

$$q^{right}_{\varphi} = I^{(1)}_{V1} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,\tau_{qy},\varphi} \tag{3.2.5}$$

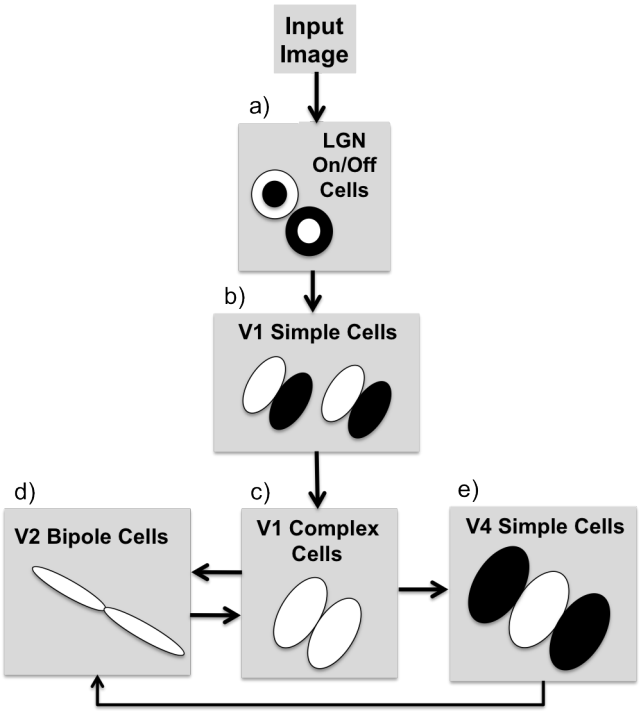

Figure 3.4.: This figure shows a simplified structure of the modified version of the texture boundary detection model.

The result of these first activations on a brain MRI image can be seen in parts *C.* and *A.* of figure 3.6. In this figure, it is also possible to see in detail the following steps of our implementation. After its first activation, V4 is activated for the last time as in equation 3.1.4, and $I_{V4}^{(2)}$ is used to modulate the activity of V2's second level of activation (top-down modulation) (See in figure 3.6 the dashed orange line that goes from *D.* towards the continuos red line that connects *A.* to *E.*). Part *E.* of figure 3.6 shows the result after V2 cells reach their highest level of activation with the intra-areal center-surround competition (equation 3.1.4). The endproduct of V2's activity finally feeds V1's second activation as stated in equation 3.1.2 by exerting the top-down modulatory effect (See *F.* in figure 3.6 and the dashed line that goes from *E.* to the red continuous line that enters *F.*).

3.3. Application to clinical imaging data

The main goal of implementing the visual attention model is its direct application on clinical MRI images in order to describe and characterize the pathological changes in them. Here we will use MRI data of patients with demyelinating lesions caused by multiple sclerosis (MS) (See Chapter 1). As explained before, the radiological diagnosis heavily depends on T2-weighted MRI images (spin echo or FLAIR) since they are able to sharply differentiate the lesions from the surrounding healthy tissue.

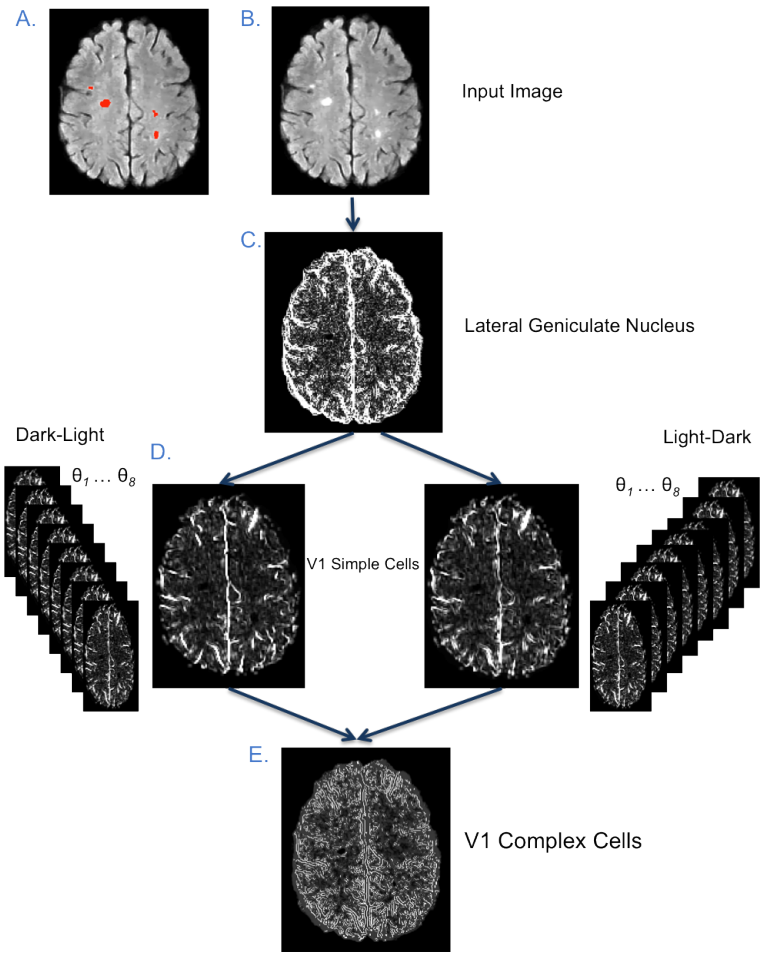

Figure 3.5.: This figure shows the bottom-up activation of LGN, V1 simple cells and V1 complex cells on an example brain MRI image with multiple sclerosis lesions

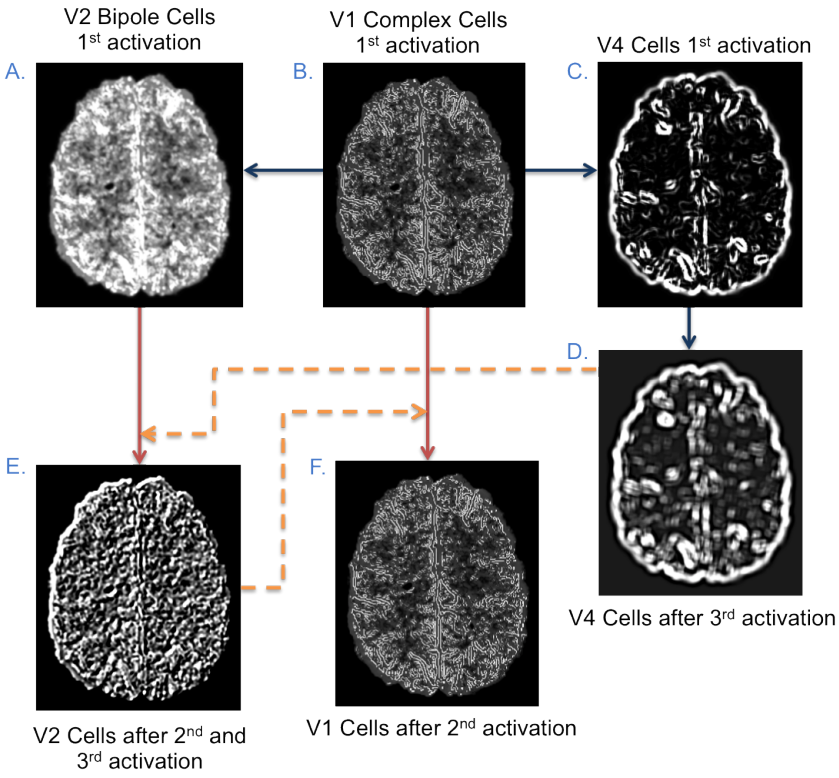

Figure 3.6.: This figure shows the bottom-up activation of V1 complex cells, V2 cells and V4 cells on an example brain MRI image with multiple sclerosis lesions. It also shows the feed-forward and feed-back interactions between these cells as they go through their different steps of activation.

A total of 23 subjects were scanned at the Magnetic Resonance Center of the San Jose Hospital in Bogota, Colombia between 2000 and 2010. These subjects were patients admitted, diagnosed and treated at the same hospital. Some patients were diagnosed with acute MS and others with recurrent chronic MS. The population mean age was 34±10.2 with 15 females and males. The images used for the analysis were acquired with a FLAIR sequence (Fluid Attenuated Inversion Recovery), that has a T2-weighted tissue contrast that suppresses the cerebrospinal fluid (CSF) signal. By such, the demyelinating lesions in the white matter are the only regions of the image with higher intensity than the rest of the brain, which makes them easier to be detected by the radiologist. The parameters of the whole-brain acquisitions were: 1.5 Tesla Siemens Vision magnetic resonance scanner (Erlangen, Germany), TE=110ms, TR=5000ms, TI=190ms, flip angle=180°, 2D axial-plane acquisition with an in-plane resolution of 256x256 and a 0.976x0.976x5mm voxel size, a total of 16 slices and spacing between slices of 2.5mm, FOV of 24.98cm.

As part of the preprocessing of the images, we used FSL's brain extraction tool BET[1] [78] to automatically skullstrip the images. After having the skull automatically removed, the masks were checked for any errors produced by the program and if there were any, they were manually fixed on Brainsuite's graphic user interface[2]. Afterwards, all the 3D brain volumes were split into separate axial slices (16 total axial slices for each subject). Only the slices that showed abnormalities were taken for further analysis. At the end we had a total of 162 axial 2D images of the brain with MS lesions. Each 2D image had a spatial dimension of 256x256.

The 162 images were smoothed with a Perona-Malik anisotropic diffusion kernel of size 0.4 and 100 iterations by using the ITK-SNAP C3D tool[3]. The same tool was also used to upsample the images with a cubic interpolation method to a spatial dimension of 512x512 pixels. Right after this step, the images were submitted to the visual texture processing module explained in the previous section.

Once the visual model was run on all the images some extra steps were performed in order to fully characterize the MS lesions of the brain. The resulting image of the second level activation of the complex V1 cells (figure 3.6 part *D.*) was substracted from the output of the second activation of V4 (See figure 3.6 part *F.*). The resulting difference image was thresholded to keep only the positive values and it was then binarized. Manual segmentations of the MS lesions were previously traced by an expert medical doctor on each of the 162 pathologic images. Finally, we computed four validation metrics, i.e. the Jaccard index, the Dice index, sensitivity and specificity, to compare the resulting relevant regions extracted from the visual computational model with the manually traced lesions.

[1] http://www.fmrib.ox.ac.uk/analysis/research/bet/
[2] http://brainsuite.loni.ucla.edu/
[3] http://www.itksnap.org/

4. Results and Conclusions

4.1. Results of Itti's model

Itti's computational model [1] is a very well known model and it is many times used as a gold standard to compare newly developed models of visual attention. Here we used an implementation of this model available on the web as an open source tool [1]. Two full brain axial slices with MS lesions and the corresponding results are shown in figures 4.1 and 4.2.

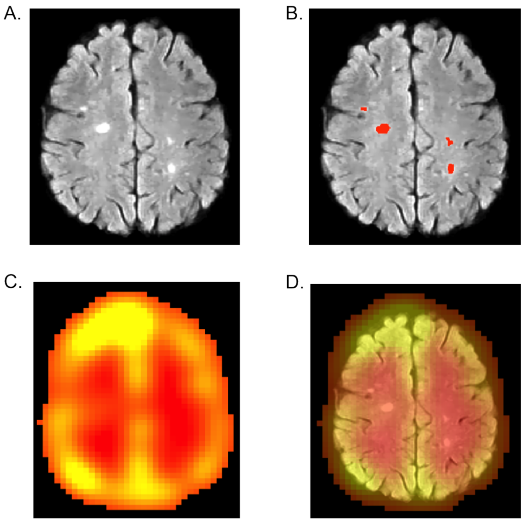

Figure 4.1.: The figure shows the input image with MS lesions (0.5 - 7cm) (A.), the manually traced mask in red overlaid on the input image (B.), the resulting saliency map (C.), and the saliency map overlaid on the MRI image with MS lesions (D.). The yellow areas are the highly salient areas, whereas the red areas are less salient ones.

Clearly, this model was not able to capture the relevant diagnostic information in these images. The saliency maps show a strong bias towards the edges of the brain tissue, rather than highlighting the MS lesions. It did not matter the size of the lesions, small or big, the model was not able to declare them as salient. This makes sense, since this model was mainly developed for natural images and applications in robotic navigation. This is a

[1] look for: http://www.vision.caltech.edu/harel/share/gbvs.php

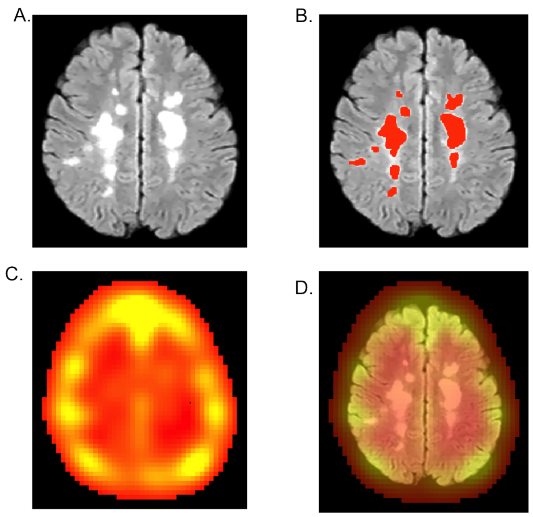

Figure 4.2.: The figure shows the input image with bigger MS lesions (7 - 10cm) (A.), the manually traced mask in red overlaid on the input image (B.), the resulting saliency map (C.), and the saliency map overlaid on the MRI image with MS lesions (D.). The yellow areas are the highly salient areas, whereas the red areas are less salient ones.

classic example that shows us how semantically different natural images can be from medical images.

4.2. Results of the proposed model

The final step of the model proposed above is the subtraction of $TP_{i\theta}^{V1}$ ("V1 second") from $CC_{i\theta}^{V4}$ ("V4 final"). The end result of this is an image that can be thresholded by eliminating all negative values. The positive values depict the salient areas for one single 2D image (see figure 4.3 for a closer look at the output on the example lesion). After running the full pipeline on the 162 images, the expert radiologist compared the manual delineations with the results of our implementation. The expert determined whether or not the proposed model was able to delineate the lesion.

The computational attention model was able to characterize MS lesions on the majority of the images. After a detailed visual inspection it could be established that the model was able to automatically delineate the lesions in 85.8% of our set of 162 images with MS lesions. Figure 4.4 shows the results for one full axial slice of the brain with small MS lesions (0.4 to 2 cm) that are distributed throughout the brain tissue in both hemispheres.

Figure 4.5 shows the results for one image with bigger MS lesions (7 to 10 cm). The first observation from these results is that the computational model was not able to pick up the smallest MS lesions. This can be seen in figure 4.4 where the two smallest lesions were not

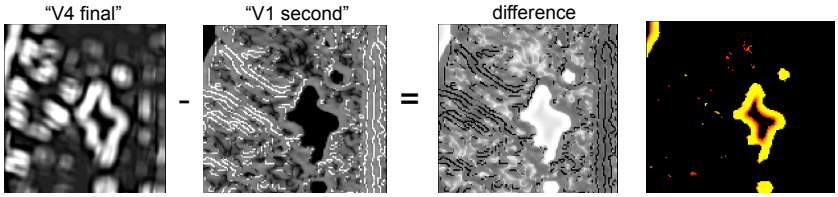

Figure 4.3.: In this figure "V4 final" is $CC_{i\theta}^{V4}$ from equation (3.1.4) and "V1 second" derived from equation (3.1.2). The colored image on the very right represents the thresholded result after subtracting the "V1 second" from "V4 final".

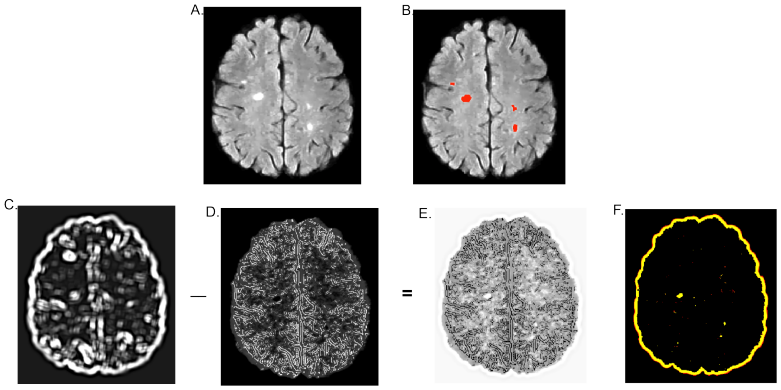

Figure 4.4.: The figure shows the input image with MS lesions (0.5 - 7cm) (A.), the manually traced mask in red overlaid on the input image (B.), the resulting image of V4 cells (C.), the resulting image of V1 cells (D.), the difference image (E.) and the thresholded version of the difference (F.)

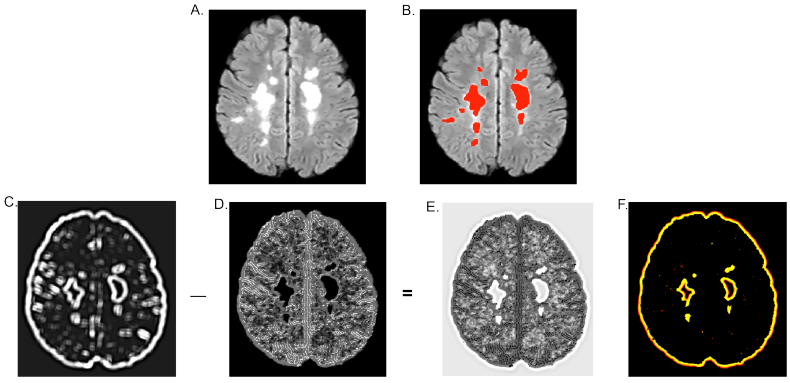

Figure 4.5.: The figure shows the input image with bigger MS lesions (7 - 10cm) (A.), the manually traced mask in red overlaid on the input image (B.), the resulting image of V4 cells (C.), the resulting image of V1 cells (D.), the difference image (E.) and the thresholded version of the difference (F.)

characterized properly by the model. Another important observation is that when the lesion is larger in average than five centimeters the algorithm will characterize it in a "ring" fashion, and would only describe the border or edge of the lesion. This can be appreciated in part F. of figure 4.5. A plausible interpretation of the results is related to the texture decomposition done by the model. The model is tuned to localize and highlight the areas of transition or change in texture, where there are contrasts of local orientations of texture elements and where there is a higher texture density [3, 79, 80]. In this case, the tissue types that are radiologically distinct are the healthy white matter and the pathological hyperintense MS lesions. After running the anisotropic diffusion filter on the images, which was a preliminary step before submitting them to the visual attention model, the images preserved the edges and the areas within these were homogenized by the filter. Thus, it is possible to argue that most of the healthy brain tissue was homogenized in terms of texture and the areas with MS lesions were also homogenized. The borders between pathological areas and the rest of the healthy brain are the areas of changing texture, the areas with most edge details (higher density) and also higher orientation contrasts. The model did not perform well in those lesions that were too small (less than 0.5 cm) maybe because there weren't enough pixels (or enough area) within them that had a homogeneous texture and not enough texture contrast was identified between the lesions and the rest of the brain tissue.

Another important observation is the fact that the visual model also highlighted the external border of the brain. According to the previous explanation, this pattern makes sense, because of the vast texture difference between the brain and the background of the image which is composed by pixels with a value of zero. It is important to note here that for the validation metrics that were run (Dice index, specificity, etc) we manually erased the brain rim spotted by the algorithm.

Table 4.1 summarizes the results of the four validation measures computed for each image. We chose as the gold standard the manually traced labels of each image. Each resulting map of visually relevant areas was compared to its corresponding manually traced label. We then computed descriptive statistics on each index separately. Interestingly, we found a very high range in the Dice, Jaccard and sensitivity indices, with minimum values starting at 0, in which case the model could not characterize the relevant visual area. Importantly, the specificity was very high for the majority of images and small standard deviation and a very small range. Although the results for the Jaccard and Dice coefficients are low, we still had a sensibility above 0.5 and a very high specificity. It is important to note here that the main goal of this study is to find a plausible explanation of how the visual system works during medical imaging diagnosis by using a known visual attention computational model for this particular task. In the future we will intent to design segmentation tools based on this approach, which may be able to be tuned to specific pathological features and a priori information.

	Jaccard	Dice	Sensitivity	Specificity
Mean	0.204	0.307	0.622	0.995
Standard deviation	0.169	0.226	0.339	0.003
Median	0.170	0.290	0.754	0.966
Minimum value	0	0	0	0.982
Maximum value	0.612	0.760	0.992	0.999

Table 4.1.: This table shows the summary of the four validation measures that were computed.

4.3. Conclusions

The first contribution of this work that I would like to highlight is the construction of a dataset of preprocessed images of MS patients. This has not been done before in Colombia and it is the first time that such a data set will be available for the research community in the field of medical image processing.

The main contribution of this work is the implementation of a biologically inspired artificial vision system for the detection of abnormalities on radiological images, specifically MS lesions on brain MRI images. Our findings show that it is possible to model the radiologist's perception and find a plausible explanation of how the human visual system works in the radiological diagnostic setting. This approach may be helpful for studying the visual system's behavior in the context of clinical radiology with functional brain imaging and eye tracking tools. Ultimately, this kind of approach will allow us to analyze artificial visual networks in a better way by introducing a priori information such as anatomical and semantic information relevant for the clinical diagnosis.

A. Appendix: SIPAIM 2013 paper

The following paper was accepted for presentation for the 9th International Symposium on Medical Information Processing and Analysis (SIPAIM) held in Mexico City November 11th, 2013 to November 14th, 2013.

Detecting multiple sclerosis lesions with a fully bioinspired visual attention model

Julio Villalon-Reina[b], Ricardo Gutierrez-Carvajal[a], Paul M. Thompson[b], Eduardo Romero-Castro[a]

[a]Computer Imaging and Medical Applications Laboratory, National University of Colombia Bogota, Bogota, Colombia;
[b]Imaging Genetics Center, Institute for Neuroimaging and Informatics, University of Southern California, Los Angeles, CA, USA

ABSTRACT

The detection, segmentation and quantification of multiple sclerosis (MS) lesions on magnetic resonance images (MRI) has been a very active field for the last two decades because of the urge to correlate these measures with the effectiveness of pharmacological treatment. A myriad of methods has been developed and most of these are non specific for the type of lesions, e.g. they do not differentiate between acute and chronic lesions. On the other hand, radiologists are able to distinguish between several stages of the disease on different types of MRI images. The main motivation of the work presented here is to computationally emulate the visual perception of the radiologist by using modeling principles of the neuronal centers along the visual system. By using this approach we were able to successfully detect multiple sclerosis lesions in brain MRI. This type of approach allows us to study and improve the analysis of brain networks by introducing a priori information.

Keywords: Multiple sclerosis, visual attention, artificial vision, magnetic resonance imaging

1. INTRODUCTION

Multiple sclerosis (MS) is the most prevalent demyelinating disease in the world. Its prevalence has been estimated to be between 2 and 25 per 100,000 habitants.[1] Its main pathological features are the destruction of the myelin of nerve fibers with relative sparing of axons, the infiltration of inflammatory cells in a perivascular distribution and lesions that are primarily located in the white matter in multiple small disseminated foci that tend to coalesce as the disease turns to its chronic phase.

Magnetic resonance imaging (MRI) is considered the main tool for diagnosis and follow-up of MS patients. This is due to the ability of different MRI sequences to depict different aspects of the disease in its different stages. T1-weighted images with contrast are able to detect acute lesions, T1-weighted images without contrast show hypointense areas of axonal damage, and T2-weighted images show acute lesions as well as shrinking chronic lesions ("T2 footprint"). Recently, magnetization transfer images have been inversely correlated with remyelination processes.[2] Although the abnormalities shown on T2-weighted images are the least specific ones, the quantification of their changes across time has been considered the standard for clinical trials.[3] The overall trend of lesions seen on T2-weighted images is to increase in number and volume over time, a phenomenon also referred to as the "T2 burden of disease". This burden is more severe in the absence of treatment and less so when there has been effective treatment. Thus, for the last 20 years there has been a rise in the development of software tools for the segmentation and quantification of MS lesions on not only T2-weighted images, but also on a combination of T1, T2 and proton density (PD) weighted images.

Most of the above mentioned methods rely on the segmentation of the lesions by using the voxel intensity. Also, the majority of the algorithms use a multichannel approach, meaning that they combine several types of MRI images, usually the T1-weighted, T2-weighted, PD, FLAIR (Fluid Attenuated Inversion Recovery, is a sequence that has a T2-weighted tissue contrast and it suppresses the cerebrospinal fluid signal) and contrast

Further author information: (Send correspondence to Eduardo Romero-Castro)
Eduardo Romero-Castro: E-mail: edromero@unal.edu.co, Telephone: 57 1 316 5491

enhanced images.[4–6] Although many studies have also used only one modality of MRI image, especially T2-weighted or PD images.[7] Another important distinction is that many methods rely on the manual input of an expert to help the segmentation process to be more accurate, which makes it a semiautomatic method,[8] whereas other approaches are absolutely automatic.[4] The other very important distinction to make is the one between supervised and unsupervised methods. The so called supervised methods rely on prior information. The prior information can be provided as either a brain probabilistic or topological atlas, or as manually pre-segmented and annotated lesions for further classification purposes.[9–11] The unsupervised methods can also be divided into two types, those that classify lesions as outliers based on a previous tissue segmentation of the brain and those that only use the lesion properties to segment them.[12,13]

After more than 20 years of research on MS lesion segmentation and quantification, no study has been published yet, to the extend of our knowledge, that uses biologically inspired algorithms to detect demyelinating pathology in brain MRI images. By biologically we mean essentially the use of physiologic principles of the visual system that have been mathematically modeled and that are worth exploring for medical image processing. The main motivation of this work is to bring one of such models to the context of brain imaging so that it is possible to detect MS pathology in MRI images and try to emulate the visual system at high levels of visual expertise like radiology.

When determining the so called "burden of the disease" most of the referred methods try to look for the total extent of the lesions. This is why the majority of the methods use multichannel approaches, since some lesions may appear independently in different types of images. One advantage of visual attention models is that they can be tuned and trained in order to detect and even describe specific types of targets in the scene. When developing such a model for medical imaging purposes, the aim is to detect and discriminate between different pathological conditions, thus it should be possible to design a system that can accurately detect a particular type of lesion. In this work we hypothesize that by using some basic principles of the visual system it is possible to develop a computational model that is able to detect MS lesions. The resulting tool will be specifically designed to detect MS lesions in T2 weighted MRI and can eventually be tuned to describe various types of lesions in the same imaging modality.

Furthermore, the combination of the current knowledge about the visual perception of radiologists with existing computational visual attention models may help disentangle many of the challenges concerning problems like accurate MS classification, staging and their relationship to therapeutics. As said before, MRI techniques are rapidly evolving and are already able to distinguish between different pathological stages of brain tissue. In this work we argue that computational visual models may be an efficient way to not only analyze pathological MRI images based on image features (luminance and texture) but also to find a suitable tool to study the neural networks involved in human visual perception.

2. METHODS

A total of 23 subjects were scanned at the Magnetic Resonance Center of the San Jose Hospital in Bogota, Colombia between 2000 and 2010. These subjects were patients admitted, diagnosed and treated at the same hospital. Some patients were diagnosed with acute MS and others with recurrent chronic MS. The images used for the analysis were acquired with the FLAIR sequence. By such, the demyelinating lesions in the white matter are the only regions of the image with higher intensity than the rest of the brain, which makes them easier to be detected by the radiologist. The parameters of the whole-brain acquisitions were: 1.5 Tesla Siemens Vision magnetic resonance scanner (Erlangen, Germany), TE=110ms, TR=5000ms, TI=190ms, flip angle=180°, 2D axial-plane acquisition with an in-plane resolution of 256x256 and a 0.976x0.976x5mm voxel size, a total of 16 slices and spacing between slices of 2.5mm, FOV of 24.98cm.

We used FSL's brain extraction tool -BET (http://www.fmrib.ox.ac.uk/analysis/research/bet/)-[14] to automatically skullstrip the images. After having the skull automatically removed, the masks were checked for any errors produced by the program and if there were any, they were manually fixed on Brainsuite's graphic user interface (http://brainsuite.loni.ucla.edu/). Afterwards, all the 3D brain volumes were split into separate axial slices (16 total axial slices for each subject).

A expert radiologist determined what slices in each of the 23 volumes were abnormal. The slices that showed abnormalities were included for further analysis yielding to a total of 162 axial 2D images of the brain with MS lesions. The radiologist also manually delineated each lesion on the selected slice. The 162 images were smoothed with a Perona-Malik anisotropic diffusion kernel of size 0.4 and 100 iterations by using the ITK-SNAP C3D tool (http://www.itksnap.org/). After these preliminary preprocessing steps we submitted the images to the visual attention model.

The model presented here is built by a set of hierarchically organized filters (also called model-cells) that represent specific groups of neurons. These filters were previously modeled by Thielscher et al.[15–17] and they resemble the different stages or steps of the the visual system's information flow. These model-cells are: the lateral geniculate nucleus (LGN), V1 simple cells, V1 complex cells, area V2 and area V4. The three latter ones, i.e. V1 complex cells, area V2 and area V4, are bi-directionally connected to each other resembling the feed-forward (bottom-up) and the recurrent (top-down) information processing and connections in the brain. V1 complex cells , V2 cells and V4 cells constitute the three higher stages of the model and each of them undergoes three successive "activation stages": 1) pooling of bottom-up activity, 2) activity modulation via feedback interaction (top-down modulation) and 3) intra-areal center-surround competition. For detailed information on the model see Thielscher et al.[15,16] We implemented a slight modification of this model, although we neither changed the core functions of the filters nor the functions that establish their different activation stages. The general structure of our model can be seen in figure 1.

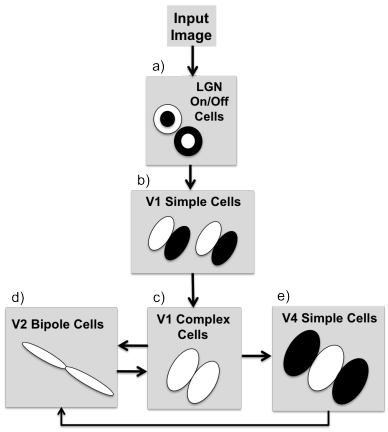

Figure 1. This figure shows a simplified structure of the modified version of the texture boundary detection model.

As is shown in part a) of figure 1, the first filter is the LGN model-cell. The LGN filter and the V1 filter (discussed below) greatly mimic the parvocellular visual pathway. This filter consists of circular center-surround receptive fields to detect local luminance transitions. If the input input is called I, the LGN is determined by:

$$\begin{aligned} x &= I * (\Lambda_{Center} - \Lambda_{Surround}) \\ x^{on} &= [x]^{+} \\ x^{off} &= [-x]^{+} \end{aligned} \tag{1}$$

I is convolved by the difference of isotropic 2D gaussian kernels represented by Λ_{Center} and $\Lambda_{Surround}$. $[x]^{+}$ and $[x]^{-}$ represent half-wave rectifications. The output of this filter with an example image is shown in figure 2.

After LGN's step the resulting ON and OFF activations (represented by x^{on} and x^{off}, respectively) are processed by the V1 filter (V1 simple cells). As it is shown in part b) of figure 1, this filter is represented by elongated ovoid ON (light) and OFF (dark) subfields and are expressed by :

$$p_{\theta}^{on/off-left} = x^{on/off} * \Lambda_{\sigma_x,\sigma_y,0,-\tau_y/2,\theta} \tag{2}$$

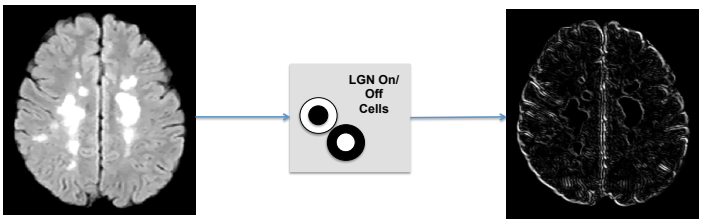

Figure 2. The resulting image on the right side shows that most borders are detected by the V1 filter.

$$p_{\theta}^{on/off-right} = x^{on/off} * \Lambda_{\sigma_x,\sigma_y,0,\tau_y/2,\theta} \tag{3}$$

In equations (2) and (3) p_θ denotes the subfields of V1 simple cells, $\Lambda_{\sigma_x,\sigma_y,0,\pm\frac{\tau_y}{2},\theta}$ are 2D anisotropic gaussian weighting functions, in which the standard deviations σ_x and σ_y define the size and shape of the subfield. θ determines the orientation of the subfield. The subfields are initially shifted perpendicular to their axis by $\pm\frac{\tau_y}{2}$ and rotated by θ ($n_{orient} = 8$), where $\theta = 0, \pi/n_{orient}, ..., (n_{orient} - 1)\pi/n_{orient}$. Thus, the analysis is split into eight maps, one for each orientation θ. The subfields $p_\theta^{on/off-left}$ and $p_\theta^{on/off-right}$ feed the first activation of V1 simple cells called S in the spatial location i and for each orientation θ. V1 simple cells exist for two polarities (ld: light-dark, dl:dark-light). The activation that is selective for light-dark polarity is:

$$S_{i\theta}^{ld} = \frac{A_s(p_{i\theta}^{on-left} + p_{i\theta}^{off-right}) + 2B_s p_{i\theta}^{on-left} p_{i\theta}^{off-right}}{A_s D_s + E_s(p_{i\theta}^{on-left} + p_{i\theta}^{off-right})} \tag{4}$$

and the activation that is selective for dark-light polarity is:

$$S_{i\theta}^{dl} = \frac{A_s(p_{i\theta}^{off-left} + p_{i\theta}^{on-right}) + 2B_s p_{i\theta}^{off-left} p_{i\theta}^{on-right}}{A_s D_s + E_s(p_{i\theta}^{off-left} + p_{i\theta}^{on-right})} \tag{5}$$

In (4) and (5) A_s, B_s, D_s, E_s are factors that control the impacts of the additive (left side of the numerator), multiplicative (right side of the numerator), and divisive components of the subfield responses. We used the same values for these parameters as in the original paper of Thielscher et al.[15,16]

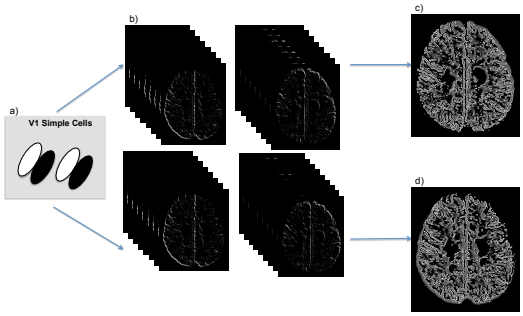

Figure 3. a) Shows a diagram of the shape of the V1's subfields. b) shows the maps derived from $p_\theta^{on/off-left}$ (equation (2)) and $p_\theta^{on/off-right}$ (equation 3). c) and d) show the end result of activations S (equations (2) and (3), respectively). Note that V1 simple cells introduce the global analysis by splitting the border detection in eight different orientations.

After the activation of V1-simple-cell filter, the filter for V1 complex cells pool the activity of two simple cells of opposite polarity ($S_{i\theta}^{ld}$ and $S_{i\theta}^{dl}$) at each position i by calculating a half-wave rectification of their difference:

$$c_{i\theta}^{V1} = A_c([S_{i\theta}^{ld} - S_{i\theta}^{dl}]^+ + [S_{i\theta}^{dl} - S_{i\theta}^{ld}]^+) \tag{6}$$

The input to the V2 filter is $c_{i\theta}^{V1}$ which is convolved by a an isotropic gaussian filter Ψ_f and by the prolated receptive field of V2 named $K^{left/right}$. K also analyzes the image globally along the eight orientations of θ.

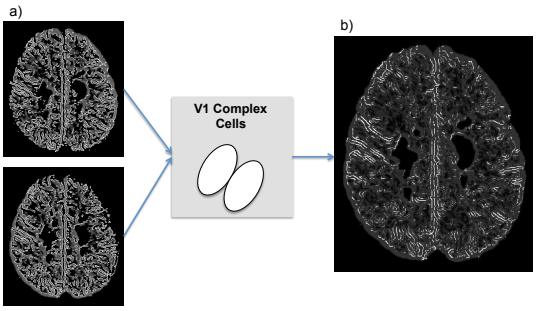

Figure 4. It is shown in the figure that the incomplete edges seen after V1 simple cells' activation are put together in order to delineate contours of the objects inside the image. a) The two images are the V1 simple cells' output and b) is the resulting image of equation (6). The same sample image is used as before.

$$f^{left} = c_{i\theta}^{V1} * \Psi_f * K^{left} \tag{7}$$

$$f^{right} = c_{i\theta}^{V1} * \Psi_f * K^{right} \tag{8}$$

Where $K_{i\theta}$ is determined by:

$$K_{i\theta}^{left} = \Lambda_{\sigma_{kx},\sigma_{ky},\tau_{kx},0,\theta}(\vec{X}_i) \times \frac{1}{1 + exp(-A_k \vec{X}_i \binom{cos\theta}{sen\theta} - B_k)} \tag{9}$$

$$K_{i\theta}^{right} = \Lambda_{\sigma_{kx},\sigma_{ky},-\tau_{kx},0,\theta}(\vec{X}_i) \times \frac{1}{1 + exp(+A_k \vec{X}_i \binom{cos\theta}{sen\theta} + B_k)} \tag{10}$$

The output image of this filter is shown is figure 5. The equation that computes V2's activation is very similar to the equation for V1 complex cells, in which the activities of opposite subfields (left and right in this case) are weighted by their additive and multiplicative interactions:

$$c_{i\theta}^{V2} = \frac{A_t(f_{i\theta}^{left} + f_{i\theta}^{right}) + 2B_t f_{i\theta}^{left} f_{i\theta}^{right}}{A_t D_t + E_t(f_{i\theta}^{left} + f_{i\theta}^{right})} \tag{11}$$

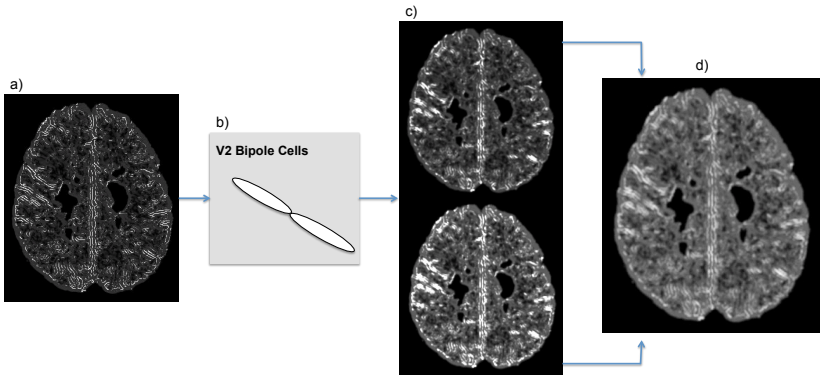

Figure 5. V2 model cells finalize the global analysis of the image by closing and delineating all edges in the image, since their receptive fields are larger and more elongated than the ones of V1 model cells. a) shows the output of V1 and d) is the final result of V2.

V1 not only passes the result of its activation ($c_{i\theta}^{V1}$) to V2 but also to V4, which is hierarchically the highest model cell of the system. This filter pools $c_{i\theta}^{V1}$ (equation (6)) into an excitatory center field called q_{φ}^{center} and measures its difference with a left and right inhibitory subfields (q_{φ}^{left} and q_{φ}^{right}).

$$q_{\varphi}^{center} = C_{V1} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,0,\varphi} \tag{12}$$

$$q_{\varphi}^{left} = C_{V1} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,-\tau_{qy},\varphi} \tag{13}$$

$$q_{\varphi}^{right} = C_{V1} * \Psi_q * \Lambda_{\sigma_{qx},\sigma_{qy},0,\tau_{qy},\varphi} \tag{14}$$

Here "$*$" denotes the convolution operator. The rotation angle of the subfields is φ and the inhibitory subfields q_{φ}^{left} and q_{φ}^{right} are shifted perpendicularly to their main axis by $\pm\tau_{qy}$. V1 and V2 filters process the information in eight different orientations ($\theta = 8$), and V4 assigns to each orientation θ another eight orientations ($\varphi = 8$). Thus, V4 suppresses most of the noise in the image wherever φ is equal to θ which can be seen in figure 6. The three subfields q_{φ} are combined in the following equation to give the output of V4.

$$c_{i\theta}^{V4} = [q_{i\theta\varphi}^{center} - Cq_{i\theta\varphi}^{left}]^{+} + [q_{i\theta\varphi}^{center} - Cq_{i\theta\varphi}^{right}]^{+} \tag{15}$$

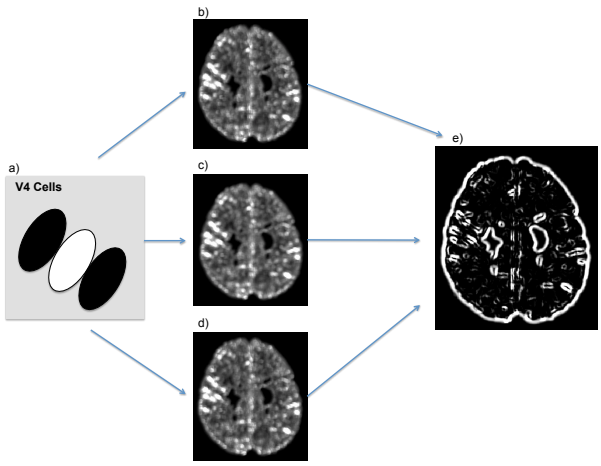

Figure 6. a) Shows a schematic diagram of the center subfield and the inhibitory left and right subfields. b), c), d) represent equations (12), (13), (14), respectively and e) shows the output of V4's first activation. At this stage the more relevant features of the original input image are the external border of the brain and some sulci borders but there is also a very sharp delineation of the MS lesion (round-shaped structures in the center of the brain).

After getting $c_{i\theta}^{V4}$ from equation (15) a "center-surround competition" follows, which consists of a border detection function that enhances the central portion of the edges and blurs or suppresses the surrounding intensities. We call V4's final output $CC_{i\theta}^{V4}$ ("V4 final") and is determined by equation (16). Figure 7 show the result of this step.

$$CC_{i\theta} = \frac{\beta_2\{C^{(V4)} * \Psi^{+} * \Lambda^{+}\}_{i\theta} - \delta_2\{C^{(V4)} * \Psi^{-} * \Lambda^{-}\}_{i\theta}}{\alpha_2 + \zeta_2\{C^{(V4)} * \Psi^{-} * \Lambda^{-}\}_{i\theta}} \tag{16}$$

After all the bottom-up information has been passed on to each of the filters, the corresponding outputs of V1, V2 and V4 interact in a top-down manner, that is to say, the higher order filters (e.g. V4) modulate the final output of lower order filters (e.g. V2). This step is determined by the following equation:

$$TP_{i\theta} = \frac{\beta_1 c_{i\theta}[1 + Ch_{i\theta}]}{\alpha_1 + \gamma_1 c_{i\theta}[1 + Ch_{i\theta}]} \tag{17}$$

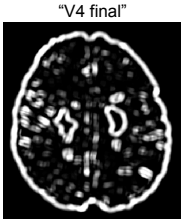

Figure 7.

In equation (17) $c_{i\theta}$ is the lower order model cell and $h_{i\theta}$ is the higher order model-cell. Thus, for the V2-V4 interaction, $c_{i\theta}^{V2}$ is modulated by $CC_{i\theta}^{V4}$ to get $TP_{i\theta}^{V2}$ which we call "V2 second" in figure 8. $TP_{i\theta}^{V2}$ then undergoes "center surround competition" as stated in (16) that yields to $CC_{i\theta}^{V2}$ and we call it "V2 final" in figure 8. Then, for the V1-V2 interaction, $CC_{i\theta}^{V2}$ modulates $c_{i\theta}^{V1}$ as in (17) to get $TP_{i\theta}^{V1}$, which we call "V1 second" in figure 9.

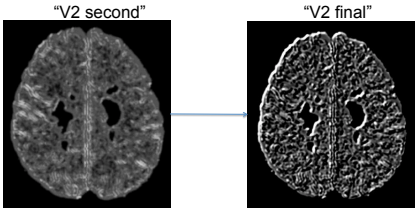

Figure 8. On the right side ("V2 final"), the borders of the MS lesions are enhanced. Notice the higher intensity around the central dark area and the homogeneous appearance of the rest of the brain tissue.

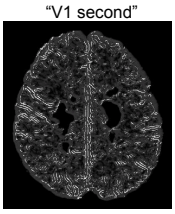

Figure 9.

The final step is to subtract $TP_{i\theta}^{V1}$ ("V1 second") from $CC_{i\theta}^{V4}$ ("V4 final"). The end result of this is an image that can be thresholded by eliminating all negative values. The positive values depict the end result of our implementation for one single 2D image (see figure 10 for a closer look at the output on the example lesion). Figures 11 and 12. show the results on two sample 2D brain images. After running the whole pipeline on the 162 images, the expert radiologist compared the manual delineations with the results of our implementation. The expert determined whether or not the proposed pipeline was able to delineate the lesion.

3. RESULTS AND DISCUSSION

We were able to automatically delineate the lesions in 85.8% of our set of 162 images with MS lesions. The first observation from these results is that the computational model is not able to pick up the smallest MS lesions. This can be seen in figure 12 where the two smallest lesions were not detected by the model. Another important observation is that when the lesion is larger in average than five centimeters the algorithm will characterize it in a "ring" fashion, and would only describe the border or edge of the lesion. This can be appreciated in part F. of figure 11. A plausible interpretation of the results is related to the texture decomposition done by the model. The model is tuned to localize and highlight the areas of transition or change in texture, where there are contrasts of local orientations of texture elements and where there is a higher texture density.[16,18,19] In this case, the tissue types that are radiologically distinct are the healthy white matter and the pathological hyperintense MS lesions. After running the anisotropic diffusion filter on the images, which is a preprocessing step, the images preserved the edges and the areas within these were homogenized by the filter. Thus, it is possible to argue

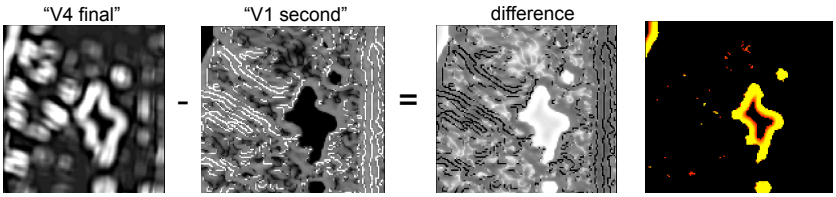

Figure 10. In this figure "V4 final" is $CC_{i\theta}^{V4}$ from equation (16) and "V1 second" derived from equation (17). The colored image on the very right represents the thresholded result after subtracting the "V1 second" from "V4 final".

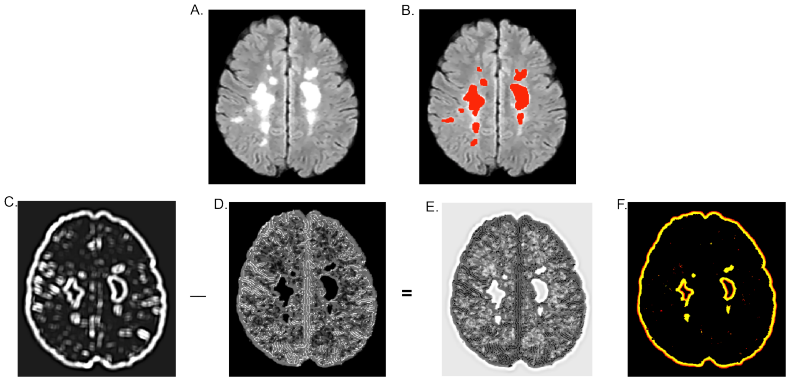

Figure 11. A) is the original input image, B) shows the expert's delineation (in red) on top of the input image, C) "V4 final" from equation (16), D) "V1 second" derived from equation (17), E) is the difference image and F) represents the thresholded subtraction.

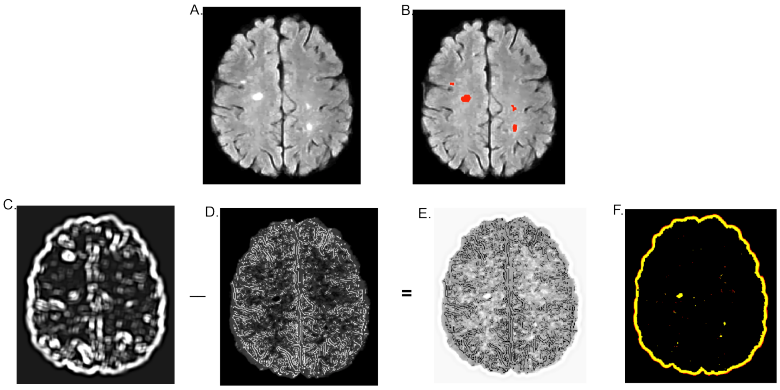

Figure 12. A) is the original input image, B) shows the expert's delineation (in red) on top of the input image, C) "V4 final" from equation (16), D) "V1 second" derived from equation (17), E) is the difference image and F) represents the thresholded subtraction. Note that the algorithm did not detect the smallest lesions that were delineated by the expert.

that most of the healthy brain tissue was homogenized in terms of texture and the areas with MS lesions were also homogenized. The borders between pathological areas and the rest of the healthy brain are the areas of changing texture, the areas with most edge details (higher density) and also higher orientation contrasts. The model did not perform well in those lesions that were too small (less than 0.5 cm) maybe because there were not enough pixels (or enough area) within them that had a homogeneous texture and not enough texture contrast was identified between the lesions and the rest of the brain tissue. In the future we will explore standardized learning methods in order to tune the model for smaller lesions detection.

We also compared the resulting detected areas against the radiologist's delineation of the MS lesions with standardized indices, i.e. Jaccard coefficient, Dice coefficient, sensitivity and specificity (Table 1). Although the results for the Jaccard and Dice coefficients are low, we still had a sensibility above 0.5 and a very high specificity. It is important to note here that the main goal of this study is to find a plausible explanation of how the visual system works during medical imaging diagnosis by using a known visual attention computational model for this particular task. In the future we will intent to design segmentation tools based on this approach, which may be able to be tuned to specific pathological features and a priori information.

	Jaccard	Dice	Sensitivity	Specificity
Mean	0.204	0.307	0.622	0.995
Standard deviation	0.169	0.226	0.339	0.003

Table 1.

Our main contribution is the implementation of a biologically inspired artificial vision system for the detection of abnormalities on radiological images, specifically MS lesions on brain MRI images. Our findings show that it is possible to model the radiologist's perception and find a plausible explanation of how the human visual system works in the radiological diagnostic setting. This approach may be helpful for studying the visual system's behavior in the context of clinical radiology with functional brain imaging and eye tracking tools. Ultimately, this kind of approach will allow us to analyze artificial visual networks in a better way by introducing a priori information such as anatomical and semantic information relevant for the clinical diagnosis.

ACKNOWLEDGMENTS

We would like to acknowledge the San Jose Hospital in Bogota, Colombia. They provided the image dataset used in this work.

REFERENCES

[1] Rosati, G., "The prevalence of multiple sclerosis in the world: an update," *Neurological Sciences* **22**(2), 117–139 (2001).

[2] Barkhof F, Bruck W, D. G. C. e. a., "Remyelinated lesions in multiple sclerosis: Magnetic resonance image appearance," *Archives of Neurology* **60**, 1073–1081 (Aug. 2003).

[3] Kappos, L., Radue, E.-W., O'Connor, P., Polman, C., Hohlfeld, R., Calabresi, P., Selmaj, K., Agoropoulou, C., Leyk, M., Zhang-Auberson, L., and Burtin, P., "A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis," *New England Journal of Medicine* **362**(5), 387–401 (2010).

[4] Zijdenbos, A., Forghani, R., and Evans, A., "Automatic "pipeline" analysis of 3-d mri data for clinical trials: application to multiple sclerosis," *Medical Imaging, IEEE Transactions on* **21**(10), 1280–1291 (2002).

[5] Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., and Suetens, P., "Automated segmentation of multiple sclerosis lesions by model outlier detection," *Medical Imaging, IEEE Transactions on* **20**(8), 677–688 (2001).

[6] Kikinis, R., Guttmann, C. R., Metcalf, D., Wells, W. M., Ettinger, G. J., Weiner, H. L., and Jolesz, F. A., "Quantitative follow-up of patients with multiple sclerosis using mri: technical aspects," *Journal of Magnetic Resonance Imaging* **9**(4), 519–530 (1999).

[7] Anbeek, P., Vincken, K., and Viergever, M., "Automated ms-lesion segmentation by k-nearest neighbor classification," *MIDAS Journal* , http://hdl.handle.net/10380/1448 (Jul. 2008).

[8] Udupa, J., Wei, L., Samarasekera, S., Miki, Y., Van Buchem, M. A., and Grossman, R. I., "Multiple sclerosis lesion quantification using fuzzy-connectedness principles," *Medical Imaging, IEEE Transactions on* **16**(5), 598–609 (1997).

[9] Wells III, W. M., Grimson, W. E. L., Kikinis, R., and Jolesz, F. A., "Adaptive segmentation of mri data," *Medical Imaging, IEEE Transactions on* **15**(4), 429–442 (1996).

[10] Warfield, S., Robatino, A., Dengler, J., Jolesz, F., and Kikinis, R., "Nonlinear registration and template driven segmentation," *Brain Warping* **4**, 67–84 (1999).

[11] Morra, J., Tu, Z., Toga, A., and Thompson, P., "Automatic segmentation of ms lesions using a contextual model for the miccai grand challenge," *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge* , 1–7 (2008).

[12] Khayati, R., Vafadust, M., Towhidkhah, F., and Nabavi, M., "Fully automatic segmentation of multiple sclerosis lesions in brain mr flair images using adaptive mixtures method and markov random field model," *Computers in Biology and Medicine* **38**(3), 379–390 (2008).

[13] Pachai, C., Zhu, Y., Grimaud, J., Hermier, M., Dromigny-Badin, A., Boudraa, A., Gimenez, G., Confavreux, C., and Froment, J., "A pyramidal approach for automatic segmentation of multiple sclerosis lesions in brain mri," *Computerized Medical Imaging and Graphics* **22**, 399–408 (Sept. 1998).

[14] Smith, S. M., "Fast robust automated brain extraction," *Human Brain Mapping* **17**(3), 143–155 (2002).

[15] Thielscher, A. and Neumann, H., "Neural mechanisms of cortico-cortical interaction in texture boundary detection: a modeling approach," *Neuroscience* **122**(4), 921 – 939 (2003).

[16] Thielscher, A. and Neumann, H., "A computational model to link psychophysics and cortical cell activation patterns in human texture processing," *Journal of Computational Neuroscience* **22**, 255–282 (2007).

[17] Thielscher, A., Klle, M., Neumann, H., Spitzer, M., and Grn, G., "Texture segmentation in human perception: A combined modeling and fmri study," *Neuroscience* **151**, 730–736 (Feb. 2008).

[18] Nothdurft, H.-C., "Feature analysis and the role of similarity in preattentive vision," *Perception & Psychophysics* **52**(4), 355–375 (1992).

[19] Nothdurft, H.-C., "Salience from feature contrast: variations with texture density," *Vision Research* **40**, 3181–3200 (Jan. 2000).

Bibliography

[1] Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Patt Anal Mach Intell **20** (1998) 1254–1259

[2] Orabona, F., Metta, G., Sandini, G.: Object-based visual attention: a model for a behaving robot. In: Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on. (2005) 89–89

[3] Thielscher, A., Neumann, H.: A computational model to link psychophysics and cortical cell activation patterns in human texture processing. Journal of Computational Neuroscience **22** (2007) 255–282 10.1007/s10827-006-0011-9.

[4] Thielscher, A., Neumann, H.: Neural mechanisms of cortico-cortical interaction in texture boundary detection: a modeling approach. Neuroscience **122**(4) (2003) 921 – 939

[5] Rosati, G.: The prevalence of multiple sclerosis in the world: an update. Neurological Sciences **22**(2) (2001) 117–139

[6] Fancy, S.P., Kotter, M.R., Harrington, E.P., Huang, J.K., Zhao, C., Rowitch, D.H., Franklin, R.J.: Overcoming remyelination failure in multiple sclerosis and other myelin disorders. Experimental Neurology **225**(1) (September 2010) 18–23

[7] Kappos, L., Radue, E.W., O'Connor, P., Polman, C., Hohlfeld, R., Calabresi, P., Selmaj, K., Agoropoulou, C., Leyk, M., Zhang-Auberson, L., Burtin, P.: A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. New England Journal of Medicine **362**(5) (2010) 387–401

[8] Miller, D.H., Khan, O.A., Sheremata, W.A., Blumhardt, L.D., Rice, G.P., Libonati, M.A., Willmer-Hulme, A.J., Dalton, C.M., Miszkiel, K.A., O'Connor, P.W.: A controlled trial of natalizumab for relapsing multiple sclerosis. New England Journal of Medicine **348**(1) (2003) 15–23

[9] Barkhof F, Brück W, D.G.C.e.a.: Remyelinated lesions in multiple sclerosis: Magnetic resonance image appearance. Archives of Neurology **60**(8) (August 2003) 1073–1081

[10] Zijdenbos, A., Forghani, R., Evans, A.: Automatic "pipeline" analysis of 3-d mri data for clinical trials: application to multiple sclerosis. Medical Imaging, IEEE Transactions on **21**(10) (2002) 1280–1291

[11] Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. Medical Imaging, IEEE Transactions on **20**(8) (2001) 677–688

[12] Kikinis, R., Guttmann, C.R., Metcalf, D., Wells, W.M., Ettinger, G.J., Weiner, H.L., Jolesz, F.A.: Quantitative follow-up of patients with multiple sclerosis using mri: technical aspects. Journal of Magnetic Resonance Imaging **9**(4) (1999) 519–530

[13] Anbeek, P., Vincken, K.L., Viergever, M.A.: Automated ms-lesion segmentation by k-nearest neighbor classification. MIDAS Journal (2008)

[14] Udupa, J., Wei, L., Samarasekera, S., Miki, Y., Van Buchem, M.A., Grossman, R.I.: Multiple sclerosis lesion quantification using fuzzy-connectedness principles. Medical Imaging, IEEE Transactions on **16**(5) (1997) 598–609

[15] Wells III, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of mri data. Medical Imaging, IEEE Transactions on **15**(4) (1996) 429–442

[16] Warfield, S., Robatino, A., Dengler, J., Jolesz, F., Kikinis, R.: Nonlinear registration and template driven segmentation. Brain Warping **4** (1999) 67–84

[17] Morra, J., Tu, Z., Toga, A., Thompson, P.: Automatic segmentation of ms lesions using a contextual model for the miccai grand challenge. Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge (2008) 1–7

[18] Khayati, R., Vafadust, M., Towhidkhah, F., Nabavi, M.: Fully automatic segmentation of multiple sclerosis lesions in brain mr flair images using adaptive mixtures method and markov random field model. Computers in Biology and Medicine **38**(3) (2008) 379–390

[19] Pachai, C., Zhu, Y., Grimaud, J., Hermier, M., Dromigny-Badin, A., Boudraa, A., Gimenez, G., Confavreux, C., Froment, J.: A pyramidal approach for automatic segmentation of multiple sclerosis lesions in brain mri. Computerized Medical Imaging and Graphics **22**(5) (September 1998) 399–408

[20] Guttmann, C.R., Kikinis, R., Anderson, M.C., Jakab, M., Warfield, S.K., Killiany, R.J., Weiner, H.L., Jolesz, F.A.: Quantitative follow-up of patients with multiple sclerosis using mri: reproducibility. Journal of Magnetic Resonance Imaging **9**(4) (1999) 509–518

[21] Ungerleider, L.G., Haxby, J.V.: 'what' and 'where' in the human brain. Curr Opinion Neurobiol **4** (1994) 157–l65

[22] Livingstone, M., Hubel, D.: Segregation of form, color, movement, and depth: anatomy, physiology and perception. Science **240** (1988) 740–749

[23] Suder, K., Worgotter, F.: The control of low-level information flow in the visual system. Rev Neurosci **11** (2000) 127–146

[24] Pasupathy, A., Connor, C.E.: Responses to contour features in macaque area v4. J Neurophysiol **82** (1999) 2490–2502

[25] Logothetis, N.K., Pauls, J., Poggio, T.: Shape representation in the inferior temporal cortex of monkeys. Curr Biol **5** (1995) 552–563

[26] Kourtzi, Z., Kanwisher, N.: Cortical regions involved in perceiving object shape. J Neurosci **20** (2000) 3310–3318

[27] Treisman, A.M., Gelade, G.: A feature integration theory of attention. Cogn Psychol **12** (1980) 97–136

[28] Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention. Annu Rev Neurosci **18** (1995) 1995

[29] O´Craven, K.M., Rosen, B.R., Kwong, K.K., Treisman, A., Savoy, R.L.: Voluntary attention modulates fmri activity in human mt-mst. Neuron **18** (1997) 591–598

[30] Treue, S., Trujillo, J.C.M.: Feature-based attention influences motion processing in macaque visual cortex. Nature **399** (1999) 575–579

[31] Roelfsema, P.R., Lamme, V.A., Spekreijse, H.: Object-based attention in the primary visual cortex of the macaque monkey. Nature **395** (1998) 376–381

[32] Valdes-Sosa, M., Bobes, M., Rodriguez, V., Pinilla, T.: Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. J Cogn Neurosci **10** (1998) 137–151

[33] O´Craven, K.M., Downing, P.E., Kanwisher, N.: fmri evidence for objects as the units of attentional selection. Nature **401** (1999) 584–587

[34] DeYoe, E.A., Brefczynski, J.A.: A physiological correlate of the 'spotlight' of visual attention. Nature Neurosci **4** (1999) 370–374

[35] Somers, D.C., Dale, A.M., Seiffert, A.E., Tootell, R.B.: Functional mri reveals spatially specific attentional modulation in human primary visual cortex. Proc Natl Acad Sci USA **96** (1999) 1663–1668

[36] Gandhi, S.P., Heeger, D.J., Boynton, G.M.: Spatial attention affects brain activity in human primary visual cortex. Proc Natl Acad Sci USA **96** (1999) 3314–3319

[37] Kanwisher, N., Downing, P., Epstein, R., Kourtzi, Z.: Functional neuroimaging of visual recognition. In: Handbook of functional neuroimaging of cognition. MIT Press (2001)

[38] Chao, L.L., Martin, A.: Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nature Neuroscience **2** (1999) 913–919

[39] Martin, A., Wiggs, C.L., Ungerleider, L.G., Haxby, J.V.: Neural correlates of category-specific knowledge. Nature **379** (1996) 649–652

[40] Forde, E.M.E.: Category specific recognition impairments: A review of important case studies and influential theories. Aphasiology **13** (1999) 169–193

[41] Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarsky, P., Gore, J.C.: Activation of the middle fusiform ''face area" increases with expertise in recognizing novel objects. Nature Neuroscience **2** (1999) 568–573

[42] Gauthier, I., William, P., Tarr, M.J., Tanaka, J.: Training 'greeble' experts: a framework for studying expert object recognition processes. Vis Res **38** (1998) 2401–2428

[43] Mervis, C.B., Rosch, E.: Categorization of natural objects. Annual Review of Psychology **32** (1981) 89–115

[44] Goffaux, V., Rossion, B.: Faces are "spatial"- holistic face perception is supported by low spatial frequencies. Journal of Experimental Psychology: Human Perception and Performance **32** (2006) 1023–1039

[45] Marr, D.: La visión: Una investigación basada en el cálculo acerca de la representación y el procesamiento humano de la información visual. Madrid Alianza (1985)

[46] Biederman, I., et al: On the information extracted from a glance at a scene. Journal of Experimental Psychology **103** (1974) 597–600

[47] Piaget, J.: La representación del mundo en el niño. Madrid Morata (1993)

[48] Friedman, A.: Framing pictures: the role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General **108** (1979) 316–355

[49] Bar, M.: Visual objects in context. Nature Reviews **5** (2004) 617–629

[50] Intraub, H.: Rapid conceptual identification of sequentially presented pictures. Journal of Experimental Psychology: Learning Memory and Cognition **10** (1981) 115–125

[51] Mance, I., Vogel, E.K.: Visual working memory. WIREs Cogn Sci **4**(2) (2013) 179–190

[52] Kastner, S., Pinsk, M.A., Weerd, P.D., Desimone, R., Ungerleider, L.G.: Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron **22** (1999) 751–761

[53] Kastner, S., Weerd, P.D., Desimone, R., Ungerleider, L.G.: Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. Science **282** (1998) 108

[54] Shulman, G.L., Corbetta, M., Buckner, R.L., Raichle, M.L., Fiez, J.A., Miezin, F.M., Petersen, S.E.: Top-down modulation of early sensory cortex. Cereb Cortex **7** (1997) 193–206

[55] Nodine, C.F., Mello-Thoms, C.: The Nature of Expertise in Radiology. In: Handbook of Medical Imaging, Volume 1. Physics and Psychophysics. SPIE Press Book (2000) 859–894

[56] Tuddenham, W.J., Calvert, W.P.: Visual search patterns in roentgen diagnosis. Radiology **76** (1961) 255–256

[57] Kundel, H.L., LaFollette, P.S.: Visual search patterns and experience with radiological images. Radiology **103** (1972) 523–528

[58] Llewellyn-Thomas, E., Lansdown, E.L.: Visual search patterns of radiologists in training. Radiology **81** (1963) 288–291

[59] Oestmann, J.W., Greene, R., Kushner, D.C., Bourgouin, P.M., Linetsky, L., Llewellyn, H.J.: Lung lesions: Correlation between viewing time and detection. Radiology **166** (1988) 451–453

[60] Nodine, C.F., Kundel, H.L., Lauver, S.C., Toto, L.C.: Nature of expertise in searching mammograms for breast masses. Academic Radiology **3** (1996) 1000–1006

[61] Kundel, H.L., Nodine, C.F.: Interpreting chest radiographs without visual search. Radiology **116** (1975) 527–532

[62] Mugglestone, M.D., Gale, A.G., Cowley, H.C., Wilson, A.R.M.: Diagnostic performance on briefly presented mammographic images. In Kundel, H.L., ed.: Image Perception. Proceedings of SPIE. Volume 2436. (1995) 106–116

[63] Lesgold, A.M., Feltovich, P.J., Glaser, R., Wang, Y.: The acquisition of perceptual diagnostic skill in radiology. Lrdc technical report pds-1, University of Pittsburgh (September 1981)

[64] Myles-Worsley, M., Johnston, W.A., Simons, M.A.: The influence of expetise on x-ray processing. Journal of Experimental Psychology: Memory and Cognition **14** (1988) 553–557

[65] Kundel, H.L.: Visual Search in Medical Images. In: Handbook of Medical Imaging, Volume 1. Physics and Psychophysics. SPIE Press Book (2000) 837–858

[66] Kundel, H.L., Nodine, C.F.: A visual concept shapes image perception. Radiology **146** (1983) 363–368

[67] Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol **4**(4) (1985) 219–27

[68] Rapantzikos, K.E., Tsapatsoulis, N.: A committee machine scheme for feature map fusion under uncertainty: the face detection case. Int J Intelligent Systems Technologies and Applications **1** (2006) 346–358

[69] Navalpakkam, V., Itti, L.: Modeling the influence of task on attention. Vision Research **45** (2005) 205–231

[70] Chernyak, D.A., Stark, L.W.: Top-down guided eye movements. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics **31** (2001) 514–521

[71] Noton, D., Stark, L.: Scanpaths in eye movements during pattern perception. Science **171** (1971) 308–311

[72] Deco, G., Rolls, E.T.: A neurodynamical cortical model of visual attention and invariant object recognition. Vision Res **44** (2004) 621–642

[73] Rybak, I.A., Gusakova, V.I., Golovan, A.V., Podladchikova, L.N., Shevtsova, N.A.: A model of attention-guided visual perception and recognition. Vision Res **38** (1998) 2387–2400

[74] Sun, Y., Fisher, R.: Object-based visual attention for computer vision. Artificial Intelligence **146** (2003) 77–123

[75] Han, J., Ngan, K., Li, M., Zhang, H.J.: Unsupervised extraction of visual attention objects in color images. Circuits and Systems for Video Technology, IEEE Transactions on **16**(1) (2006) 141–145

[76] Petroudi, S., Brady, M.: Breast density segmentation using texture. In Astley, S., Brady, M., Rose, C., Zwiggelaar, R., eds.: Digital Mammography. Volume 4046 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2006) 609–615

[77] Marr, D., Hildreth, E.: Theory of edge detection. Proceedings of the Royal Society of London. Series B. Biological Sciences **207**(1167) (1980) 187–217

[78] Smith, S.M.: Fast robust automated brain extraction. Human Brain Mapping **17**(3) (2002) 143–155

[79] Nothdurft, H.C.: Feature analysis and the role of similarity in preattentive vision. **52**(4) (1992) 355–375–

[80] Nothdurft, H.C.: Salience from feature contrast: variations with texture density. Vision Research **40**(23) (January 2000) 3181–3200

[81] Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory. COLT '92, New York, NY, USA, ACM (1992) 144–152

[82] Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology **2** (2011) 27:1–27:27 Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[83] Chen, P.H., Lin, C.J., Schölkopf, B.: A tutorial on v-support vector machines. Applied Stochastic Models in Business and Industry **21**(2) (2005) 111–136

[84] Haxby, J.V., Gobbini, M., Furey, M., Ishai, A., Schouten, J.L., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science **293** (2001) 2425–2430

[85] Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V.: Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci USA **96** (1999) 9379–9384

[86] Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision **42** (2001) 145–175

[87] Torralba, A., Oliva, A., Catelhano, M.S., Henderson, J.M.: Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search. Psychological Review **113** (2006) 766–786

[88] Vapnik, V.: An overview of statistical learning theory. Neural Networks, IEEE Transactions on **10**(5) (sep 1999) 988 –999

[89] Wei, X., Warfield, S.K., Zou, K.H., Wu, Y., Li, X., Guimond, A., Mugler, J.P., Benson, R.R., Wolfson, L., Weiner, H.L., Guttmann, C.R.: Quantitative analysis of mri signal abnormalities of brain white matter with high reproducibility and accuracy. J. Magn. Reson. Imaging **15**(2) (2002) 203–209

[90] Wu, Y., Warfield, S.K., Tan, I.L., Wells, W.M., Meier, D.S., van Schijndel, R.A., Barkhof, F., Guttmann, C.R.: Automated segmentation of multiple sclerosis lesion subtypes with multichannel mri. NeuroImage **32**(3) (2006) 1205–1215