INDIAN STATISTICAL INSTITUTE, KOLKATA

MASTER'S THESIS

Bio-inspired networks: From DoG to CNN

Author:
Rahul Kumar Ojha

Supervisor:
Dr. Kuntal Ghosh

A dissertation submitted in partial fulfillment of the requirements for the degree of

Masters Of Technology

in

Computer Science

July, 2018

Dedicated to my mother.

Declaration of Authorship

I, Rahul Kumar Ojha, declare that this thesis titled, "Bio-inspired networks: From DoG to CNN" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this institution.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this institution or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

CERTIFICATE

This is to certify that the dissertation entitled "Bio-inspired networks: From DoG to CNN" submitted by Rahul Kumar Ojhato Indian Statistical Institute, Kolkata, in partial fulfillment for the award of the degree of Masters Of Technology in Computer Science is a bonafide record of work carried out by him under my supervision and guidance. The dissertation has fulfilled all the requirements as per the regulations of this institute and, in my opinion, has reached the standard needed for submission.

Dr. Kuntal Ghosh
Associate Professor
Machine Intelligence Unit (MIU)
Indian Statistical Institute, Kolkata

List of publications

1. R.K. Ojha, D. Mazumdar, K. Ghosh: Understanding the Geometry of visual space through Muller-Lyer Illusion. Accepted for poster presentation at Spatial Cognition 2018, Tuebingen, Germany, September 5-8, 2018.
2. S. Maitra, R.K. Ojha, K. Ghosh: Impact of Convolutional Neural Network Input Parameters on Classification Performance. Submitted to 4TH IEEE International Conference on Convergence of Technology(I2CT), Mangalore, India, October 27-28, 2018.

Abstract

There has been a quest to understand the structure and functionality of the brain, in general, and the human visual system, in particular, since centuries. One reason that the quest prevails among engineers too, is because of the belief that if we can understand the human brain and visual system, it could help us design and develop models for similar tasks with similar high performance and accuracy. This, on the other hand, may perhaps also aid to arrive at a unified computational model for vision and may open new avenues in artificial vision. The persual of the quest has thus resulted in a good amount of research in the field of bio-inspired models, especially inspired from mammalian vision system. In this work, a study of several computational models for different levels of vision has been performed and its application in varying domains have been explored.

The models, studied here, revolve around the central theme of David Marr's organisation of hierarchy of vision as an information processing system and the bio-inspired models for each level of it. A major part of the work revolves around mid-level visual representation of an image and study of biologically inspired models like difference of Gaussian, which is also a computational model for the response of retinal ganglion cells, and those of Lateral Geniculate Nucleus (LGN). The basic DoG along with its variants like Extended Difference of Gaussian (EDoG), Oriented Difference of Gaussian (ODoG) and Dynamic EDoG have also been explored as possible unified approaches to low-level and mid-level vision.

For instance, the EDoG is already known for modeling of the response of the non-Classical Receptive Field (nCRF) of retinal ganglion cells with an extra Gaussian in comparison to DoG that leads to better edge map(Ghosh, Sarkar, and Bhaumik, 2005b). This dissertation work consist of three approaches to explore bio-inspired models for the three levels of Marr's hierarchy for vision. The first part of the work is dedicated to using the EDoG for understanding how geometry around an object effects its perceived size. In this part, structural and geometric information present in an image is used to find out how size depends on shape using the EDoG

model. To do this a geometric illusion, namely The Muller Lyer Illusion (MLI) has been used for study. More precisely, EDoG has been used to understand how the lengths of the lines in the illusion are perceived. Further, the role played by geometry of and around an object in perceiving its length is observed. To do this, the relation between a critical parameter of the illusion (angle between the wings) and the induced illusion is also investigated. The results obtained from computational model have been compared with the experimental results for verification. This shows that the EDoG can be a plausible model of mid-level vision, beyond edge representation.

The second part of this work is devoted to study a modified adaptive version of EDoG model (Wei, Wang, and Lai, 2012). By using the EDoG model with reverse control mechanism of vision, the brightness intensity information contained in an image has been used to give a good mid-level representation. It is shown that the application of the modified version namely dynamic nCRF (the original static version, as already mentioned in the previous paragraph, was envisaged to provide more meaningful edge information as compared to DoG), provides visually meaningful segmentation of images. Further, a modified version of this algorithm is proposed which produces a more unified mid-level representation of image by computing the image segments as well as the edge map at the same time. So both the static and the dynmaic versions of EDoG are potential candidates for mid-level representation in that they can make groupings, estimate outlines (that can be significant for motion) as well as size from shape.

As the final investigation into the higher level vision in Marr's hierarchy, more complex models are considered in the last part of the work. In this regard, we have considered convolution neural network (CNN) and deep learning which are one of the most extensively researched topics at the moment. The deep CNN has a biological motivation in the context of brain-like computing, and has been found to perform exceptionally well in pattern recognition in many complex vision problems. It needs to be noted that the Receptive field modeling through spatial filters like DoG (low-level) or EDoG (mid-level), described in the earlier chapters, are also convolutional networks in a primitive form. Hence the investigation

of deep CNN in the context of high-level vision, is logical and relevant. To relate the same to Marr's approach, an extensive study of the relevance of input parameters of a deep CNN, like number of convolutional layers, convolution kernel size, number of filters in one layer etc. with the classification accuracy and training time of the system is performed considering one example of a five-class problem using a well-known color fundus dataset for classifying diabetic retinopathy. The results are further analyzed in light of the biological structures of the human vision system.

Acknowledgements

I would first like to thank my thesis advisor Dr. Kuntal Ghosh. The door to his office was always open whenever I ran into a trouble spot or had a question about my work or writing. He consistently allowed this thesis to be my own work, but steered me in the right direction whenever he thought I needed it.

I would also like to thank Dr. Sanjit Maitra from ISI North East Center for his guidance towards my work. Without his passionate participation and input, the some portions of this work could not have been successfully completed. I would like to express my gratitude to Anjan Da, Joginder and my friends for keeping a positive workplace.

Finally, I must express my deepest gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of research and writing of this thesis. This accomplishment would not have been possible without them.

Thank you.
Rahul Kumar Ojha
M Tech, Computer Science
Indian Statistical Institute, Kolkata

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In everyday life we come across a number of varying visual experiences ranging from looking at a traffic scene on a sunny day to staring at stars in a clear sky, but still we are able to move and make a clear picture of the world seamlessly everywhere. In our day to day life, we also come across a lot of inconsistent visual inputs. Still we make up some information out of these. Some of these apparently inconsistent figures fall in the category of illusions. This apparently effortless and smooth experience is a product of millions of years of evolution which has given us a complex enough model to have a look at the world. This complex system is better known to us as visual system. Mammalian visual system as ours is capable of doing a number of complex jobs from edge detection to object recognition and more.

Biologically, the mammalian visual system consists of a number of parts. Spatially it can be seen as two broad components, viz retinal and cortical. Amongst both of these pre-cortical retinal cell structure modeling has been well studied topic from computational modeling perspective. It is easier to study biologically as shown by (Curcio et al., 1990). The pre-cortical system is of great importance because of two facts: Firstly, this part serves as a preprocessing step in human visual system, understanding of which will help us in improvement of a number of computer vision problems, Secondly a better modeling of pre-cortical regions will imply a better understanding of the human visual system. The various existing models, developed for the retinal cell modeling are also to be studied. This opens up the possibility of finding usefulness of the models in different vision related problems. Cortical area for vision in contrast are very complex to look at. For this reason, only the abstract nature

could be understood rather than observing at the network level.

The human race has been in a quest to find the organization and functionality of the visual system from computational modeling aspect. The construction of a unified model for vision would prove to be a seminal advancement. (Marr, 1982) gave one of the most renowned model for vision. According to him the vision system is an information processing system with different level of hierarchy for different representation of image. Each level/representation can achieve different type of work. He proposed the concept of 3 level image representation where the crude image is represented in different forms to achieve different jobs. He named the layers as Primal sketch, 2.5D sketch and 3D sketch. A primal sketch of an image is based on feature extraction from fundamental components of the scene. It includes edges, regions, etc. This representation is regarded as the low-level representation. Whereas a 2.5D sketch of the scene includes more informations like segments, texture information, etc. It is broadly regarded as the mid-level image representation. The final one is the 3 D model. In a 3D map of an image, the image is visualized in a continuous, 3-dimensional map. It can be used to do the higher level tasks like of object detection, face recognition, depth perception and so on.

In this work we aspire to gain a better understanding of the computational models corresponding to each level of the hierarchy to understand the vision system better. We explore what are the bio inspired models for each layer of representation in Marr's hierarchy for vision and their applicability in different tasks. A review of existing models for the first two level of representation is presented in the next section along with some topics which are required to understand the models.

1.2 Existing Models and concepts

1.2.1 Retinal biology

The retinal cells present in our visual system is arranged in form of a five layer neural network consisting of different types of cells having different functionality as shown in Figure 1.1. This cell's nature and the network has been studies for long as shown in (Davson and

Perkins, 2018). The bottom most layer comprises of a series of rods and cones which are photo receptors, working as a transducer to convert the light energy to chemical and electrical potentials.

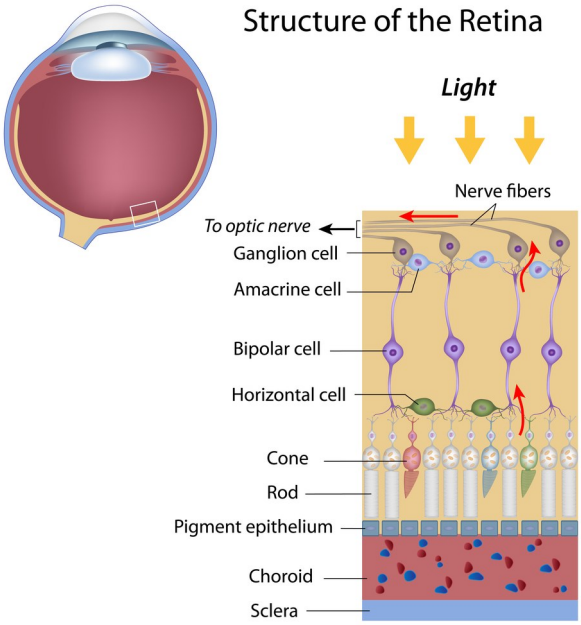


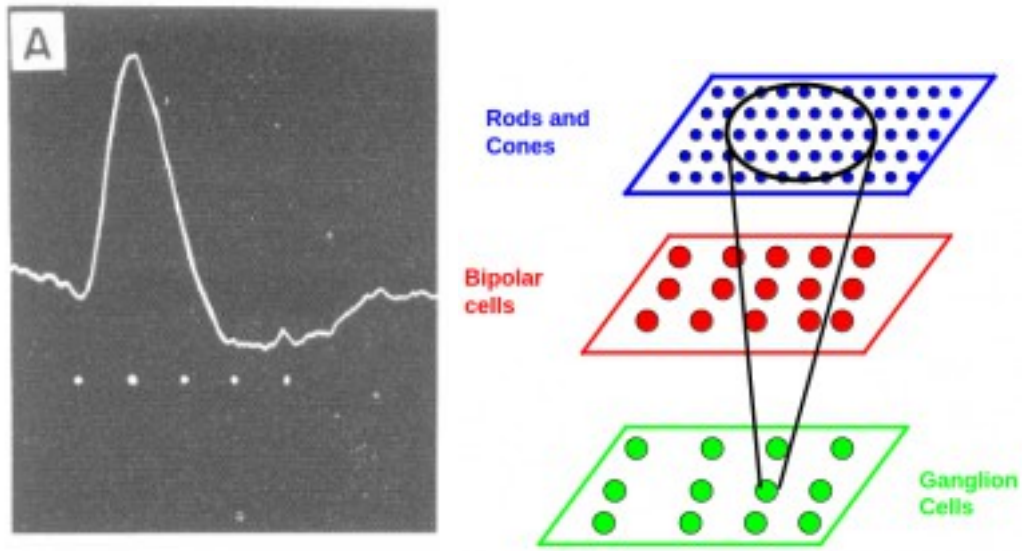
FIGURE 1.1: Neural net at retina.

The captured information then moves to horizontal cells from where it moves up to bipolar cells. Bipolar cells forward the processed information to amacrine cells then to ganglion cell. Retinal ganglion cell in mammals has been studied well in particular. A number of models have been developed for the mentioned portion of retinal neural net. Some of those are discussed in this section. The dissertation work mainly includes some modification and application of the discussed models.

1.2.2 Concept of Receptive field

(Kuffler, 1953) was one of the first ones to have captured the firing rate of a single retinal ganglion cell in mammals against an input stimulus of light.

Later on multiple attempts were made in same direction to find the neuron firing rate of the retinal ganglion cell in order to model the response. It is to be noted that ganglion cells have a certain base firing rate which is considered as the baseline. Any change in the baseline firing rate is recorded, Figure 1.2a represents the same. It



(A) Firing rate of a single ganglion cell as found by (Kuffler, 1953).

(B) Receptive field from photo receptors to ganglion cells.

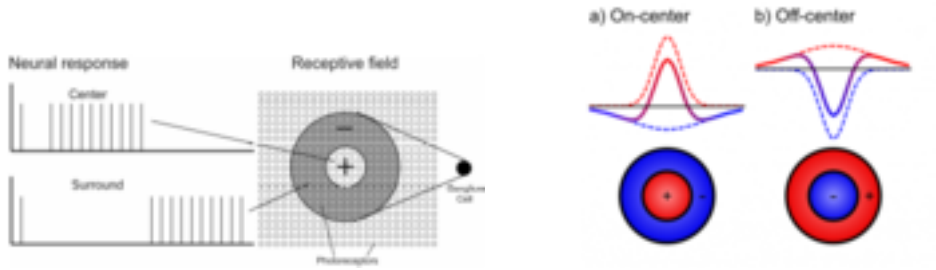
FIGURE 1.2: Receptive field and firing rate.

has been observed that the activation of a particular neuron in the higher layer depends upon the activity of a certain region of the input layer only. This region of interest is called **receptive field** of this neuron for a particular neuron. Following this analogy, lightning a certain region of the retina indicates a change in firing rate of a ganglion cell. According to (Hubel and Wiesel, 1962) *"receptive fields of cells at one level of the visual system are formed from input by cells at a lower level of the visual system."* In this way, small, simple receptive fields could be combined to form large, complex receptive fields. Hence for a particular ganglion cell, a cone shaped volume is created such that only values of elements in that cone matters to this ganglion cell and any element outside this has no effect in change of activity of this neuron. This is represented in Figure 1.2b. It shows the mapping containing ganglion cells,bipolar cells along with rods and cones.

1.2.3 Difference of Gaussians (DoG)

The response of a neuron is not same throughout its receptive field, for the same intensity input. Instead has maximum effect in a circular region near center and is less significant in a region outside this circle. This circular region, for which the neuronal activity is

maximum is known as **center** and the outside region is known as the **surround**. This model containing them together is known as **Classical Receptive Field(CRF)** or center-surround receptive field.



(A) Change in firing rate at different regions. (B) DoG model figure.

FIGURE 1.3: DoG region-wise firing rate.

Further, the nature of a ganglion cell and its response also varies according to (Davson and Perkins, 2018). In particular the nature of change in rate of firing of ganglion cell is not same for all ganglion cells. In a certain type of ganglion cells a stimulus in center region increases the rate of firing, which keeps dropping as one moves away from the center. It keeps dropping until a stimulus in the surround region decreases from the baseline firing rate. This indicates firing rate to be inversely proportional to distance from center of the center region. This phenomenon of going below baseline is called **inhibition** of surround. This inhibition is shown in Figure1.3a. In the figure the positive marked region means positive change in firing rate where a negative region marks suppression or inhibition. In analogy, a bipolar cell receives direct inputs from many retinal receptor cells (RCs) which forms the CRF, center of the bipolar cell. The bipolar cell receives indirect input from more RCs through horizontal cells. This further form the CRF surround of the bipolar cell. (Hartline, Wagner, and Ratliff, 1956) showed a modeling approach in which the center region's activity can be estimated as a Gaussian function of positive peak direction and the surround region's inhibition activity is estimated as a negative Gaussian function . (Hartline, Wagner, and Ratliff, 1956) hence modeled the activity of such type of ganglion cells as a combination of two Gaussian of different variance and same mean. This model for ganglion cell is known as **Difference of Gaussian(DoG)** model. These particular type of cells are known as **On-center Off-surround** ganglion cells, named because of their nature of firing. From physiological studied

it has been found out that this is followed up in hierarchy. Alternatively another type of ganglion cells are there where the pattern of firing rate is exactly opposite to that of On-center Off-surround cells which are known as **Off-center On-surround** ganglion cells as the name suggests the central part is modeled as a negative Gaussian and surround is modeled as a positive Gaussian.

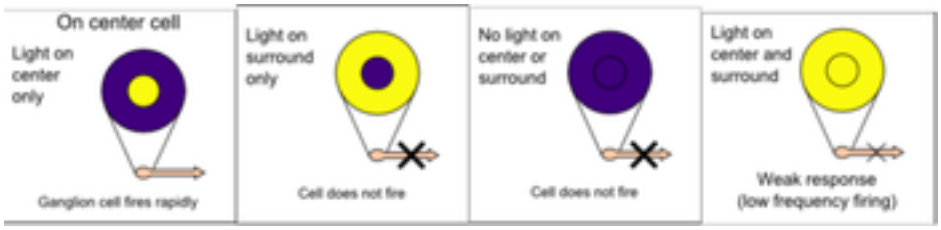


FIGURE 1.4: On-center response

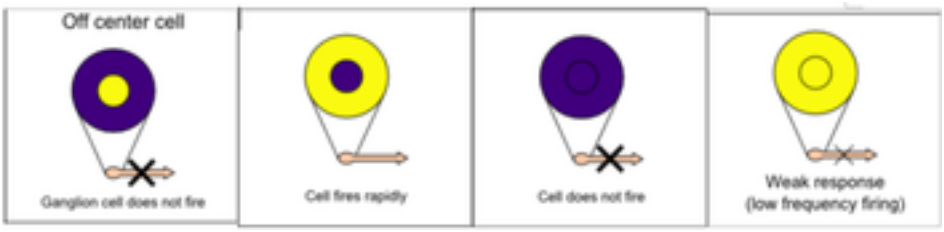


FIGURE 1.5: Off-Center On-Surround response

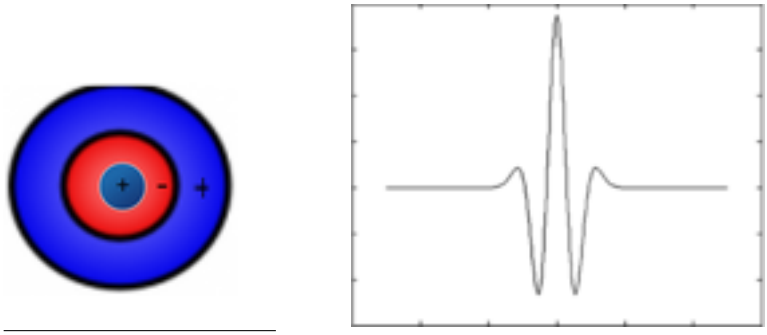
The nature of both the cells is shown in Figure 1.4 and 1.5. It can be seen that a stimulus in both the region cancels effect of center and surround and hence very less change in firing rate is observed. The mathematical formulation of DoG function can be given as follows.

$$DoG(\sigma_1, \sigma_2) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x)^2/2\sigma_1^2} - \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x)^2/2\sigma_2^2}$$

where $\sigma 1$ and $\sigma 2$ are standard deviation of center and surround Gaussian and $\sigma 1 < \sigma 2$. The resultant Gaussian is shown in Figure 1.3b.

This model of CRF has spatial summation properties. This enables this model to detect boundaries of images. DoG has been well estimated to LoG which is a edge detector in (Marr and Hildreth, 1980). Hence, it could be recognized as a model for the first level of the primal sketch as suggested by (Marr, 1982). This is one of the most fundamental model fir the primal sketch since it can perform

the edge detection as well as can help in other primal sketch activity too like indicating regions.



(A) Change in firing rate at different regions in nCRF.

(B) EDoG model response.

FIGURE 1.6: ECRF region and graph.

1.2.4 Extended Difference of Gaussians (EDoG)

DoG was long considered the ultimate model for ganglion cell because of its excellent ability to match the curve for neuron firing rate change as well as to be able to explain a number of phenomenons like illusions, brightness perception, etc. Yet it failed to explain many subtle effects. At the same time, an apparently small but relevant questions remained about existence of a faint firing rate change when both the center and surround is presented with stimuli. The answer to the question was found later by (McIlwain, 1966) and others. They found that there exists an area outside the center and surround of the classical model where presentation of stimuli when alone does not creates much of neuronal activity but when presented together with the center region produces a higher response then without it. This suggests the existence of a region outside the classical region which contributes to the neuronal activity of the associated ganglion cell. This region outside the surround is called the **extended surround**. This model is known as **non-Classical Receptive Field (nCRF)**. The model is also called **Extended Classical Receptive Field(ECRF)**since it contains an extended Gaussian into the classical model. It is also known as **Extended Difference of Gaussian(EDoG)**, since it contains an extra Gaussian from the DoG model. Hence the terms nCRF, ECRF and EDoG has been used in literature and in this dissertation also. nCRF has been modeled as a combination of three zero-mean Gaussian at three different scales

by (Ghosh, Sarkar, and Bhaumik, 2005b). This is equivalent to a bi-harmonic or bi-Laplacian of a Gaussian filter. (Ghosh, Sarkar, and Bhaumik, 2005a) and (Ghosh, Sarkar, and Bhaumik, 2006) used a linear three-Gaussian function to model nCRFs. They further seek to explain certain brightness contrast illusions. It is also to be noted that the contrast, orientation, and direction of motion of the stimulus stimulating the surround affects the suppressive effect. The regions are shown in Figure 1.6a. It is to be noted that the earlier models for DoG have considered an antagonistic behavior of the regions. But now there are three components so each of the Gaussian has be given a weight for its contribution to firing rate change.

This mathematical formulation is often also known as the **Extended Difference of Gaussian(EDoG)** which contains an extra Gaussian term as follows.

$$EDoG(\sigma_1,\sigma_2,\sigma_3) = a_1 * \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x)^2/2\sigma_1^2} - a_2 * \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x)^2/2\sigma_2^2}$$
$$+ a_3 * \frac{1}{\sigma_3\sqrt{2\pi}} e^{-(x)^2/2\sigma_3^2}$$

where $\sigma_1, \sigma_2 and \sigma_3$ are standard deviation of the three Gaussians and $a_1, a_2 and a_3$ are the empirical constants determining the relative contribution of the different regions towards the activity of the neuron. The overall accumulated response is shown in Figure 1.6b.

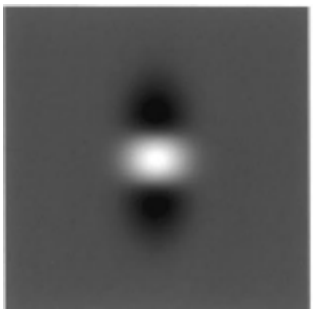


FIGURE 1.7: An oriented DoG at angle of orientation 0^0

This model of Extended DoG along with its variants have been shown to perform better segmentation and other higher levels task with the help of an additional Gaussian. Hence, it could be considered as the model for the 2.5D sketch. The processed imaged shows

the idea of textures too. Its ability to provide a better mid-level representation is explored more in Chapter 2 and Chapter 3.

1.2.5 Oriented Difference of Gaussian(ODoG)

Not only the pre-cortical but early locations in the visual cortex like V1 has also been an entity of equal or perhaps more interest. In a landmark paper (Hubel and Wiesel, 1959) showed the presence of another type of cells which unlike the retinal ganglion cells respond to only stimuli in an oriented manner rather than responding to symmetric stimuli. In other words these only respond to lines of different orientation and that too with some particular spatial frequency. So a certain type of cell only responds to a line oriented at a certain angle. Each set of cells are tuned to a spatial orientation and frequency.

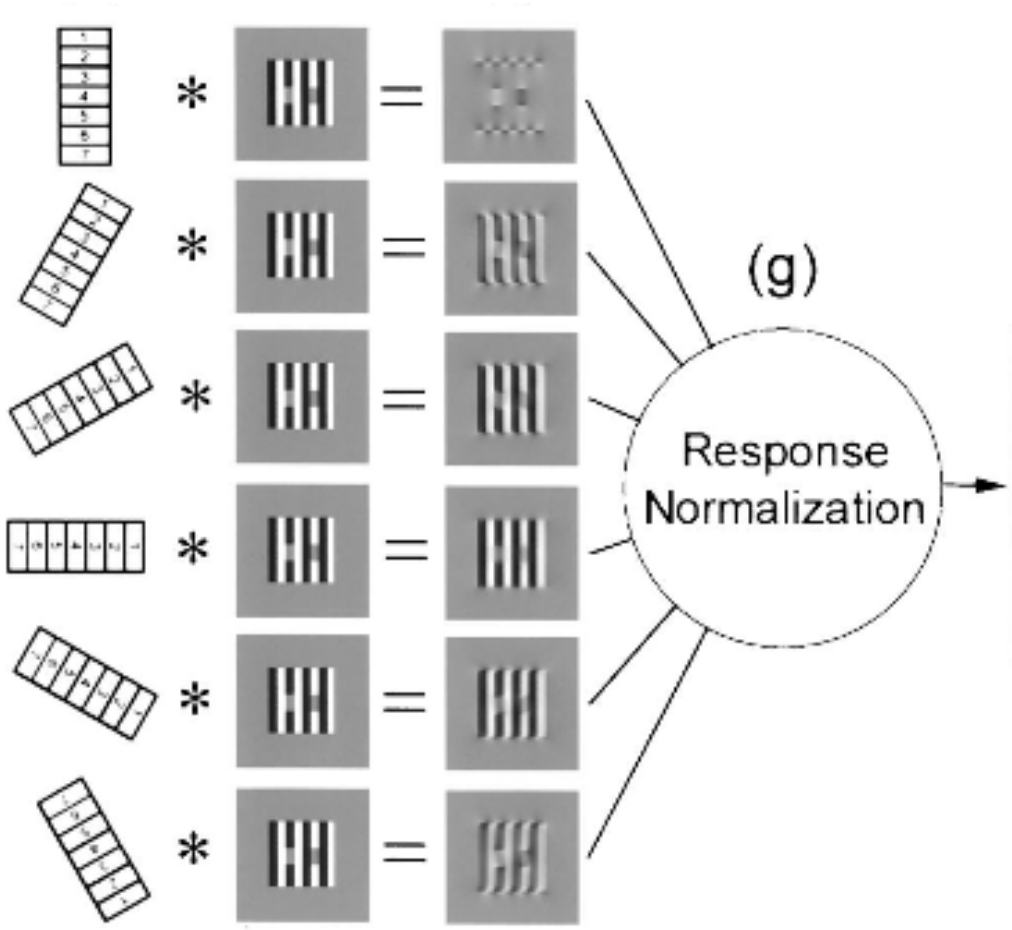


FIGURE 1.8: (Blakeslee and McCourt, 1999) model for Simple cells

(Hubel and Wiesel, 1959) named it as **simple cells**. A number of ways have been taken to model these cells like by (Daugman, 1985), who modeled the response as a Gabor filter. This modeling gives a nice estimate but lacks the biological corresponding. This Gabor model is not analogous to the anatomical structure of the visual system. It doesn't take the LGN into account and uses the 2D image as it is projected on the retina. This indicated the presence of an oriented Gaussian which is shown in at an orientation of 0^0 in Figure 1.7 Later on this oriented behavior of simple cells were modeled in (Blakeslee and McCourt, 1999). (Blakeslee and McCourt, 1999) represents this behavior of the simple cells as a combination of a number of DoG filters which are oriented different directions and also of different scales. They use a total of 42 filters which includes all combinations of seven scales and six orientations. The orientations are from $0^0 to 180^0$ in a distance of 30^0 each. This modeling is much more biologically plausible as there are clear evidence of occurrence of DoG in pre-cortical regions and a later orientation effect can be achieved biologically too. They further apply a response normalization process and accumulate the results to produce a single response map. The representation is shown in Figure 1.8. The process resembles that of applying a filter bank of an image to get an image pyramid.

1.3 Thesis Layout

This thesis consists of five chapters. The first introductory chapter, which is the present one, includes a literature review of a number of bio inspired models for different layers of Marr's hierarchy. The second chapter explores the aspects of using geometric and shape information to understand how an image is perceived using the ECRF model. It presents a proposed method using ECRF to explain the Muller Lyer illusion and attempt to find the relevance of the attributes of the illusion with the induced illusion to understand cognition of space through image. This work has been accepted for poster presentation in an international conference Spatial Cognition 2018. The third chapter contains work on the modified version of the ECRF model for the purpose of better mid-level representation by using the brightness information in the crude image. This work

is presently under consideration for preparing a manuscript for future submission. The fourth chapter explores about the models for third level of vision in Marr's hierarchy. It presents more computational aspect of modeling by understanding impact of input parameters of deep convolutional neural network on classification performance. This has already been submitted for peer review in an IEEE conference (I2CT 2018). The final chapter depicts the conclusions and planned future work from each of the chapters.

Chapter 2

Explaining illusions by estimating size from shape

2.1 Abstract

A number of models have been proposed and studied for the human visual system in order to perform usual tasks like object recognition, face recognition, etc. A class of them investigates the structural information present in the image to get the size information of objects in the image. This class of algorithms tend to find the size of an object from geometry in the image. We try to study one such approach in this chapter. In order to study this approach we use geometric illusions. To understand how the size of an object is affected from its geometry, geometric illusions becomes one of the most interesting problems to look at. Further, the fact that an ability to explain the exceptions make a model more closer to reality should also to be appreciated. The work proposed is to explain how length is perceived in the Muller Lyer illusion(MLI) (a geometric illusion) using a bio-inspired model namely, nCRF(non Classical Receptive Field) discussed in Chapter 1. The nCRF model which has been used is also considered a model for the 2.5D representation of an image by (Marr, 1982). The proposed work further investigates the relevance of a crucial parameter forming the Muller Lyer illusion, the angle between converging or diverging arrows, with the percentage of induced illusion for a better understanding of space cognition. The percentage of induced illusion, experienced by humans found experimentally, is compared with the percentage of induced illusion, indicated by the proposed method, with respect to change in the angle between wings.

2.2 Introduction

An image when presented to a complex image processing or computer vision system always converts it to an intermediate representation for better understanding and feature extraction which can be used by higher modules to do complex tasks like object recognition, motion detection and so on. Our brain similarly has been believed to have a certain representation of the crude image that it gets from the retina which is then forwarded to higher regions to do the complex tasks as discussed in Chapter 1 by (Marr, 1982). A considerable understanding of some bio-inspired model for finding these inherent representation of image will certainly help in both computer vision and cognitive science community.In particular the understanding of type of inherent representation of an image in human visual system and its modeling will certainly help in a number of improvements in the efficiency of computer image processing.

Most of the methods studied in this regard in the community to understand visual system explores only about the brightness information content in the image not the geometrical aspects present in the image. The alternative way to explore about an image is the geometrical information contained in the image. The geometric information infers the shape and size of an object present in an image. In order to extract this size information from geometry of an image the nCRF model has been used which is discussed in Chapter 1. In order to test the method proposed for estimating size we use geometric illusion as a test case. Illusions have been an area of much interest for a long time. It shows some of the most critical structural and functional aspects of human visual system. Geometric illusions, in particular, are worth exploring for the task.

A number of approaches have been taken in past to explore about geometric illusions using models corresponding to lower level model in Marr's hierarchy like DoG by (Mandal, 2016).At the same time the 2nd tier model like nCRF have been used to explain brightness illusions by (Ghosh, Sarkar, and Bhaumik, 2006) where they used nCRF to explain the White effects. In contrast in this work an attempt has been made to explain the geometric illusion MLI with the help of a bio inspired model nCRF which has been seen as a model for the 2nd layer in Marr's hierarchy. The attempt to explain MLI and hence understanding how size is perceived using nCRF is also

very important because either the models which have been used to explain MLI are not as versatile as nCRF or the versatile models nCRF have been used to explain brightness illusions. Discussed next are the MLI in detail followed by a discussion on perceptual field understanding of which is required for the methodology section.

2.2.1 Muller Lyer illusion

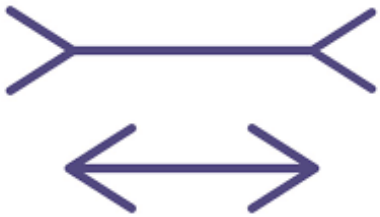


FIGURE 2.1: The Muller Lyer illusion.

The Muller Lyer Illusion (MLI) is a classical geometrical illusion of size. In this illusion, perceived line length is decreased by arrowheads and increased by arrow tails as shown by (Day and Knuth, 1981). It is a classic case of image inducing misjudgment in length of two lines of same length. The Muller Lyer illusion is shown in Figure 2.1. Some of its variations are shown in Figure 2.2.

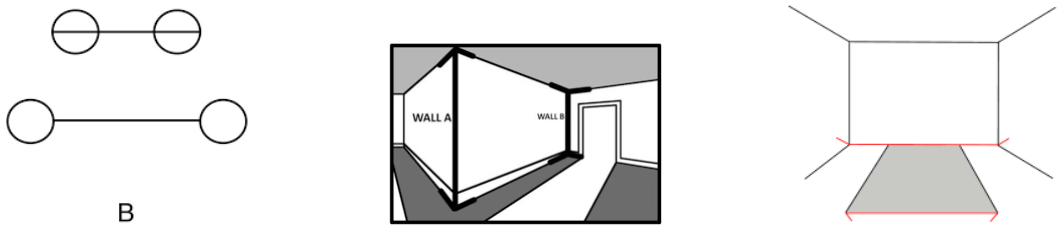


FIGURE 2.2: Variations of Muller Lyer illusion.

It is a classic example of a case where the geometry of and near a point in space influences the perceived position of the point in the space. Many theories have been put forward to explain the MLI. Still there is ongoing debate for the source of the MLI.

2.2.2 Perceptual field

The concept of the perceptual field and its use for mathematical analysis for visual illusion was first given by (Kawabata, 1976). They emphasized that the visual information captured by retina while transferred to cortical areas for vision goes through a lot of neuronal activity and modifications. Since the activation of higher level neurons is nothing but superposition of a number of receptive fields the image as received by cortical regions is also different than the original image. This representation is referred to as perceptual field by them and analysis of this perceptual field will certainly give properties of the perceptual image.

2.3 Methodology

2.3.1 Convolution filter: nCRF

In Chapter 1 idea of the non Classical Receptive field and its mathematical formulation is discussed in detail. In Section 2.2 its has been discussed how the nCRF model has been used to explain different brightness illusions. As stated in Section 2.1 nCRF has been used in this case too for convolving with the illusion image and getting a representation of it, followed by getting a contour plot which indicates the neuronal activity(discussed in next section) to get the perceived length of the lines. To recall the nCRF filters mathematical formulation is given as follows:

$$EDoG(\sigma_1,\sigma_2,\sigma_3) = a_1 * \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x)^2/2\sigma_1^2} - a_2 * \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x)^2/2\sigma_2^2}$$
$$+ a_3 * \frac{1}{\sigma_3\sqrt{2\pi}} e^{-(x)^2/2\sigma_3^2}$$

where $\sigma_1, \sigma_2 and \sigma_3$ are standard deviation of the three Gaussian and $a_1, a_2 and a_3$ are the empirical constants determining the relative contribution of the different regions towards the activity of neuron. these coefficients $a_1, a_2 and a_3$ are drawn from the same relation used in Section 3.3.1.

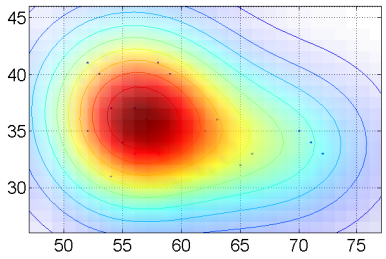


FIGURE 2.3: Sample contour plot

2.3.2 Contour Plot

In Chapter 1 a number of models representing the retinal neural activity has been discussed. Since the models represent the activity of retinal neural net, it then must indicate the changes in retinal image while transferring to cortical area. Hence using these models, we can get the perceptual field and perceptual image of an image. So it can be concluded from above arguments that the perceptual field of an image could be found through the application of models discussed in Chapter 1.

In order to view this perceptual field we have used a contour plot through Python programming language. A contour plot is a graphical technique for representing a 3-dimensional surface on a 2D colored plane. It is done so by plotting constant z slices, called contours, on a 2-dimensional format. For a given of z, lines are drawn for connecting the (x, y) coordinates where that z value occurs. The contour plot can be seen as an alternative to a 3-D surface plot. A sample contour plot is shown in Figure 2.3.

The python library matplotlib has been used to create the contour plot of the converted image or perceptual field. To summarize, the main approach for explaining the illusion used proposed in this work is to convolute the Muller Lyer illusion image with a nCRF filter with a particular set of parameters which are biologically most plausible. This convoluted image is then subjected to method described in Section 2.3.2 to find a contour plot of the filtered image. With the help of this contour plot the perceived end points and perceived distance of lines.

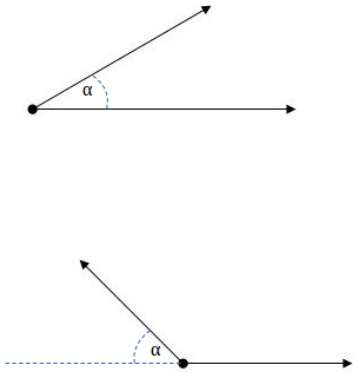


FIGURE 2.4: Angle between wings.

2.3.3 Relevance of angle

The angle between the wings (α) is an essential parameter for MLI. It has been shown via experimental data by (Restle and Decker, 1977) that angle between the wings of converging or diverging arrow affects the amount of induced illusion. To understand the relevance of the angle between the converging or diverging arrowheads of lines in MLI. For the sake of better understanding the angles between the axis and the diverging or converging wing of the arrowhead is named α. This angle is kept same for both diverging and converging wing for the sake of consistency in induced illusion. The angle is shown in Figure 2.4.

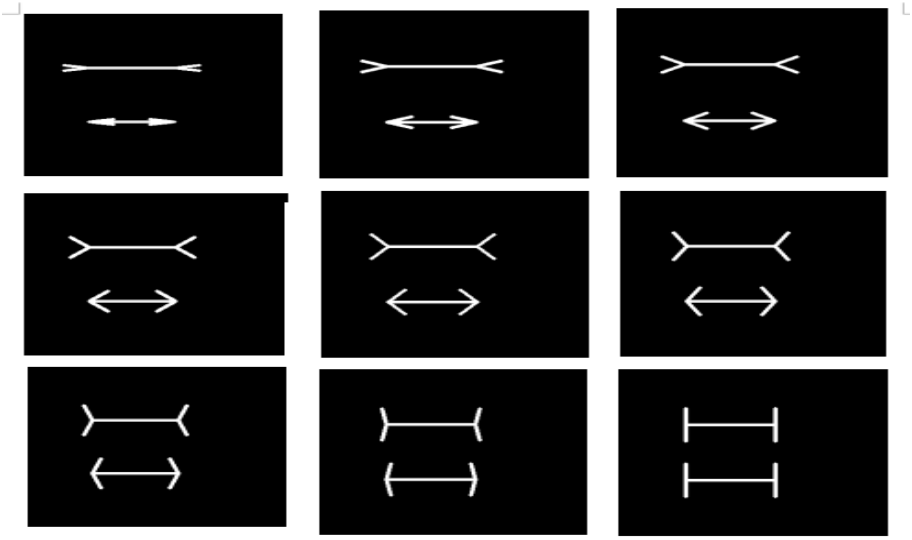


FIGURE 2.5: Angle between wings.

α is varied from 2^0 to 90^0. All the images are subjected to the

methodology discussed in Section 2.3.2. The perceived length of both the lines is found for all the varied angle images. Some of the images are shown in Figure 2.5.

2.4 Result

2.4.1 Original Muller Lyer illusion

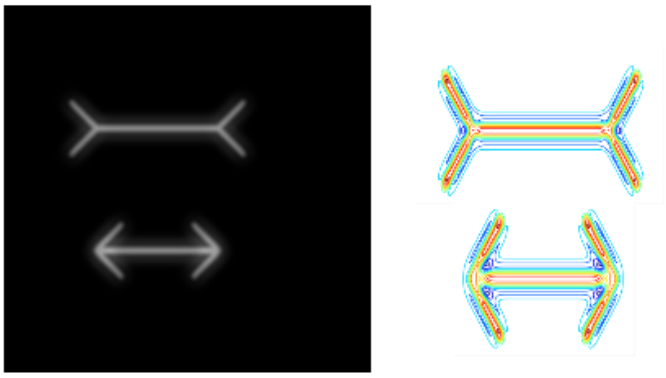


FIGURE 2.6: Contour plot of filtered MLI image.

The original MLI is subjected to the method proposed. On the convolution of the illusion image with nCRF filter produced an image shown in Figure 2.6, seeming not very different from original image except for a small blur. Further when this resultant image is subjected to contour plot, the result is shown in Figure 2.6 and the perceptual field or contour plot is shown in Figure 2.7. The contour plots show the presence of circular artifacts at the junction of arrow wings and the lines.

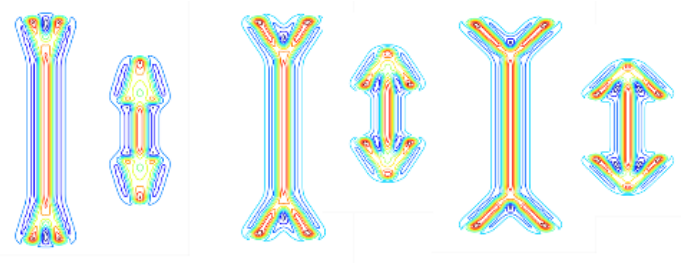


FIGURE 2.7: Contour plot of filtered image.

These artifacts are marked by high valued(red colored) circular shapes. This in terms of neuronal activity represents increase in firing rate and hence taken as the *perceived end points* in the perceptual

field. Once these endpoints are identified via the artifacts. The coordinates for the exact endpoints could be identified manually by roughly locating the center for the circular artifact. In order to find the *perceived distance* the geometric distance between these two endpoints for a line is considered. Once the perceived distance is found, the percentage illusion could be found by comparing the perceived distances for both converging and diverging arrowed lines. This perceived distance is calculated for an MLI image with length of line being 150 pixels and lengths of the wings being 30 pixels. The proposed method is applied on the MLI image a number of times and the average of the perceived distance is taken for the finding induced illusion percentage. 10 trials on the image give an average perceived length of 113.5 pixels for diverging arrowed line where as for converging arrowed line the average was found as 96 pixels. In order to find the induced illusion percentage the following formula is used:

$$Illusion percentage = \frac{Length of diverging arrowed line - Length of converging arrowed line}{Length of original line}$$

where length is Perceived length of a line. In this case the induced illusion percentage comes our as 16.5%. It is a encouraging result since it shows a significant perceived illusion length and also very close to experimental data for adult human as by (Restle and Decker, 1977).

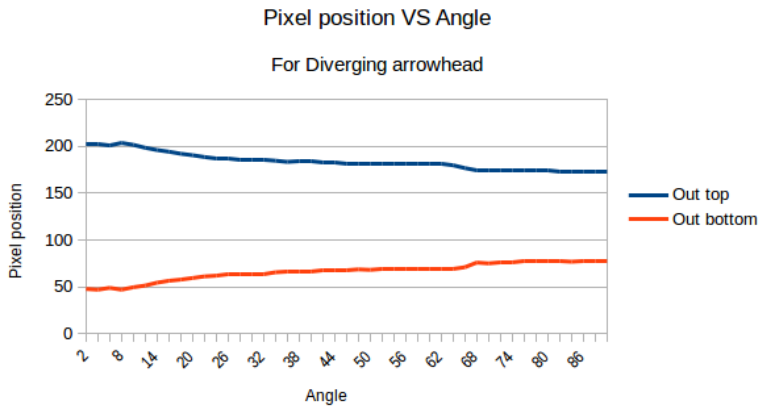


FIGURE 2.8: Plot for perceived pixel position vs angle for diverging arrowhead line.

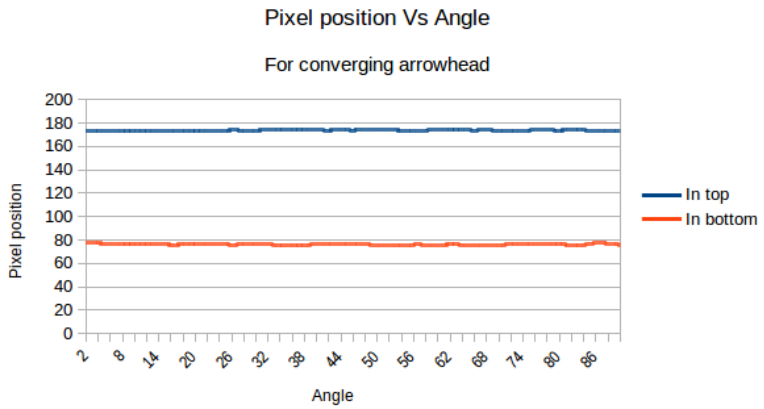


FIGURE 2.9: Plot for perceived pixel position vs angle for converging arrowhead line.

2.4.2 Relevance of angle between wings

The MLI configuration at 45 different angles from 0^0 *to* 90^0 is used to find the induced illusion. Each of the images has the same configuration as for the original image i.e. length of the line is 100 pixels and length of wings is 30 pixels. The images, unlike previous case, are subjected to nCRF filter and then contour plot only once. The plots for the perceived pixel location of the end points of both the lines are given in Figure 2.8 and Figure 2.9.

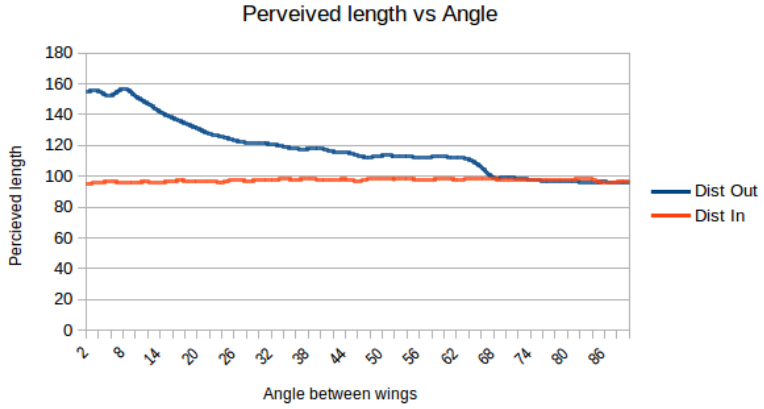
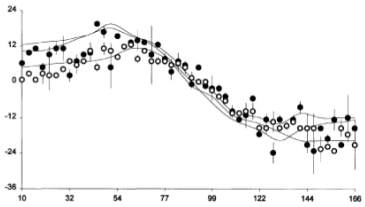


FIGURE 2.10: Plot for perceived length vs angle for both line.

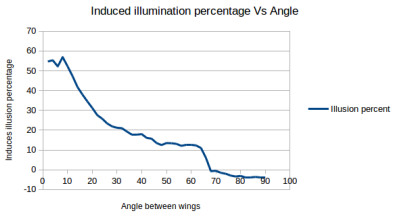
In the figures top indicated the perceived pixel position of higher endpoint colored in blue for both the plots whereas the bottom indicates the perceived pixel position of lower endpoint colored in orange. The plots show that the perceived pixel position for diverging arrowed line decreases with the increases in the arrow angle whereas doesn't change much for converging arrowhead. The plot

for perceived lengths for both the lines is given in Figure 2.10. In the plots the blue line represents perceived length for the diverging arrowhead line and orange line shows the same for converging arrowhead line.

2.4.3 Comparison with experimental data



(A) Plot for percent illusion vs angle between wings from (Bulatov, Bertulis, and Mickienė, 1997).



(B) Plot for percent illusion vs angle between wings from our approach.

FIGURE 2.11: Comparison of percent illusion from experimental and computational data.

The plot for experimental data is given in Figure 2.11a. The induced percent illusion from the computational approach proposed is shown in Figure 2.11b. It could be seen that the nature of the curves in both the plots is similar. There exist an initial rising pattern in both the plots succeeded by a sharp decrease in the induced percentage illusion.

Chapter 3

Dynamic Extended Classical Receptive field and its applications

3.1 Abstract

In this chapter the ability of ECRF based models for giving a better mid level representation of an image and hence it's candidature for being a model for 2nd level in Marr's hierarchy is explored. In order to show that ECRF as a model can be a good candidate for modeling 2nd level in Marr's hierarchy in this chapter, one of its modification is shown to be giving better mid level representation of an image. The variance of the applied model in this chapter is in the use of brightness information present in the image along with a top down approach in vision. The variant model known as the dynamic ECRF was originally proposed by (Wei, Wang, and Lai, 2012). In this work their model is understood and implemented to see how it is giving a better representing by segmentation process. Further in this work a variant of the original work by (Wei, Wang, and Lai, 2012) has been developed which produces both a mid level representation of the image and also give an edge map of the original image from a single algorithm. This work further gives support o the candidature of ECRF as a model for both 1st level and 2nd level in Marr's hierarchy.

In the work by (Wei, Wang, and Lai, 2012), they show that the human visual and attention system is not only a bottom up process i.e. from retina to higher parts of the brain but has also been found to be a top down process i.e. from higher parts of brain to retina. Various modeling techniques have been given for this assimilation of both top down and bottom up method of vision and attention. Adaptive receptive field size and dynamic non classical receptive field is one of the models proposed by them for this process. In their work the size of each receptive field is calculated dynamically as a part of the top down process. They have shown it to provide better mid

level representation of an image by showing that it is doing better for segmentation tasks.

3.2 Introduction

According to (Marr, 1982) the human visual system can be considered as an information processing system where an image in its crude for is converted to a number of representations of different tasks to be performed easily. It includes conversion to an intermediate representation for better understanding and feature extraction which can be used by higher modules to do complex tasks like object recognition, motion detection and so on.

In this regard it is important to see that the information content of a crude image is broadly explored in two ways, one of it is to explore the brightness or intensity content of the image and provide a better representation of image so that more complex tasks could be easily performed. The other way is to use the structural and geometric information present in the crude image. This alternative way helps cognition of the shape and size of an object and space around it.This chapter deals with use of inherent brightness and intensity information in the image to find a better mid level image representation, performing better at segmentation. Where as model based on the alternative way is discussed in Chapter 2.

Now to understand how the information content could be used for getting better representation of image the significant questions which are to be asked are

- What is the nature of inherent representation of image?
- What are the biological mechanisms for it?
- How to incorporate these mechanisms in computational model?

Each of the succeeding subsection will discuss about each of the question.

3.2.1 Mid level image representation

A single pixel has very limited information content. It can noway be used for finding any semantic information content. A number of

neighboring pixel only can give a semantic information about the image. This happens via the representation of an image on different levels as suggested by (Marr, 1982) in this book. He suggests and is believed widely that image is represented on mainly 3 different levels. First level in the physical representation is similar to the retinal image, containing zero crossings edges bars, etc. (Marr, 1982) names it as the primal sketch. Second level is that of a grain size of the block It contains more semantic information than the first level image. It mainly contains information about color intensity, shows broad segments and textures. In this regard this second level of image representation is referred to as **mid-level image representation**. It has a very important role in the in-line processes of vision. The mid level image representation is more of an abstract but still general in nature. It is also to be noticed that this mid level representation is only possible with integration of pixels on image based features itself. The third level representation contains information and semantic clues up to the level of object. It contains the 3d and background information too.

3.2.2 Biological mechanism for mid level image representation

The human visual system as discussed must also attain a mid level visual representation in order to provide the higher regions so that these can have a final representation of image and can do the complex tasks like object detection, etc. The main components of the human vision system in the retinal region are the cells and arrays in the ganglion layer of the retina, the lateral geniculate nucleus (LGN), and the primary visual cortex (V1). The constant change in the input via the retina due to fast change in surrounding is a very important point to note. This indicates a self-adaptive mechanism to cope up with external varying stimuli. The concept of the receptive field is already discussed in the Section 1.2.2 the existing models for it like CRF and nCRF has also been discussed in the subsequent sections. (Wei, Wang, and Lai, 2012) believe that the nCRF based model which explain the neuronal activity in retinal ganglion cells and LGN give the neural basis for the integration of features in a local region. They also believe that this is done using an intermediate size scale. Of almost all the models that have been studied it

was (Wei, Wang, and Lai, 2012) who first incorporated the top down methodology.

3.2.3 Reverse control mechanism

Research shows that size of the ganglion RF depends on changes in brightness, background, the duration of stimuli, the speed of moving objects, and so on (Li, 1997). Since there are evidences from electro-physiological, anatomical, and morphological perspectives that modulation coming from higher cortical layers is of a functional importance. A model must include top-down feedback. This is done using a reverse control mechanism. It is a top down process. The more physiological evidence includes interplexiform cells, exerting backward control over horizontal cells as well as bipolar cells to change the size of the CRF surround. Similarly the mesencephalon exercise backward control over interplexiform cells and amacrine cells for changing the size of the CRF center through centrifugal fibers. The cortico-geniculate pathway also has top-down modulations (Webb et al., 2002). These give enough evidence for presence of a reverse control mechanism in pre cortical regions in human visual system for the purpose of mid level image representation.

3.2.4 Self Adaptive receptive field

By now it is evident from biological and physiological findings that there is a need for a reverse control mechanism for finding mid level image representations in our visual system. But the exact mechanisms for it is still under question.As discussed in section 3.2.3 that the size of receptive field changes with the external factors. Broadly speaking to locate and represent borders and details of objects smaller RFs are used, while to represent the regional area of an image larger RFs are used . To look at a portion of image in detail, the portion require a smaller RF size where as to have a look at a larger object as a whole it needs larger RF size too. A detailed region must be having high spatial frequency where as the recognition of an object of higher scale as a whole requires the opposite analogy. This dynamic behavior can only be incorporated via its correspondence with retinal visual cells. One of the ways suggested

by (Wei, Wang, and Lai, 2012) was to adjust the size of the receptive field for each ganglion cell computation dynamically. This dynamic adjustment needed to be both task specific including the top down method as well as depends upon the receptive field central pixel neighborhood in image. The receptive field size change in this dynamic method is named the **self-adaptive receptive field**

3.3 Methodology

The methodology used in this work is mostly drawn from (Wei, Wang, and Lai, 2012) as described in the Section 3.1. In order to find a mid level representation of image with the advantage of the top down approach the size of the filter to be convolved with is adjusted dynamically. This mechanism of the dynamic changes in the RFs of GCs must have some important criteria like the rules determining how to find RF size must apply to any image, the adjustment in RF size should be performed automatically and mechanically. This dynamic adjustment is done using the self adaptive size of the receptive fields. (Wei, Wang, and Lai, 2012) have done it by observing surround at a certain GC's receptive field central pixel. If the surrounding is homogeneous then, the size of RF must be increased and do it until it hits an inhomogeneous region where there is a considerable difference in neighboring pixels. To quote (Wei, Wang, and Lai, 2012), *we propose the following mechanism: a GC first detects the properties of adjacent small areas in an image, then, if it finds the properties are similar, it extends its RF to integrate and represent them; in contrast, if the properties are dissimilar, the RF shrinks, enabling the properties to be distinguished and represented separately.* They name it as dynamic nCRF. In section 3.3.1 the pseudo code for finding the size of the receptive field for every GC is given along with finding the convoluted result when convolved with nCRF filter of dynamically found receptive field size.

3.3.1 Dynamic nCRF

To recall from 1.2.4 the nCRF filter could be found by using the following expression:

$$EDoG(\sigma_1, \sigma_2, \sigma_3) = a_1 * \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x)^2/2\sigma_1^2} - a_2 * \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x)^2/2\sigma_2^2}$$

$$+ a_3 * \frac{1}{\sigma_3\sqrt{2\pi}} e^{-(x)^2/2\sigma_3^2}$$

where $\sigma_1 \sigma_2 and \sigma_3$ are standard deviation of the three Gaussian and $a_1, a_2 and a_3$ are the empirical constants determining the relative contribution of the different regions towards the activity of neuron. Here the parameters are considered fixed for this implementation. The values of parameters $a_1, a_2 and a_3$ are found using the relations established in (Wei, Wang, and Lai, 2012) with $\sigma_1, \sigma_2 and \sigma_3$ as follows

$$a_1 = \frac{100}{\sigma_1^2}$$

$$a_2 = a_1 * \left(\frac{\sigma_1}{\sigma_2}\right)^2$$

and

$$a_3 = 0.5 * a_1 * \left(\frac{\sigma_1}{\sigma_2}\right)^2$$

Algorithm 1 Simplified algorithm for Dynamic Adjustment of RF size

```
Set initial parameters min, max scale initial size for RF
for each central pixel for convolution do
    size ← init_size
    GC_val ← image convolved with nCRF filter of size
    Calculate GC_val-GC_val_small and GC_val-GC_val_large
    if GC_val-GC_val_small differs from GC_val-GC_val_large then
        for size = init_size − 1 : min_size do
            GC_val ← GC_val_size
            GC_val_small ← GC_val_size − 1
            if GC_val not differs much from GC_val_small then break
    else
        for size = init_size + 1 : max_size do
            GC_val ← GC_val_size
            GC_val_large ← GC_val_size + 1
            if GC_val differs much from GC_val_small then break
    intensity_value_this_pixel ← GC_val
```

The algorithm presented is a simplified form of the algorithm by (Wei, Wang, and Lai, 2012). In the algorithm for each pixel which is a candidate for center pixel during convolution process an iterative process is repeated (line 2). Initially a minimum and maximum scale for the algorithm is set which acts as the smallest and largest filter size after adaptation of the receptive field(line 1). In this case min size is set as 10 pixels whereas max size is set as min of length or breath of image input. An initial size of the filter is set which hold for each iterative process. Initially the value of image convolved with nCRF filter of initial size is stored in GC_val(line 4). Now for each of the candidate central pixels its neighborhood is checked for similarity starting from initial size of the filter. To do this GC_val_small and GC_val_large is calculated which are convolved result between image and nCRF filter of size *size*-1 and *size*+1 respectively(line 5). GC_val_small and GC_val_large are compared to check the region of the filter(line 6). If the convolved value of initial size filter differs much from convolved value of smaller size filter then the region is not homogeneous (line 6) other wise the region is homogeneous(line 11). In the earlier case the region(filter) size is decrease and increased in later case. While decreasing the size of the filter gradually to get homogeneous region until min_size (line 7) the convoluted value of current region is found(line 8) and convoluted value of smaller region is found(line 9). Comparison between these two values GC_val and GC_val_small is done after each pass to find whether homogeneous region achieved(line 10). Similarly for the other case while increasing the size of the region gradually(line 12) convoluted value of current and larger region is kept in GC_val and GC_val_large(line 13,14). These are compared to check if current region still remains homogeneous after each pass(line 15).Finally when desired region has been found it is stored as intensity value of the pixel(line 16).A representation of the iterative process is given in Figure 3.1.

3.3.2 Adaptive receptive field size as edge detector

The algorithm for dynamic nCRF provides an image as filtered with dynamically adjusted receptive field size. Most of the models which have been studied till now on dynamic nCRF (Wei, 2016) (Wei, Dai,

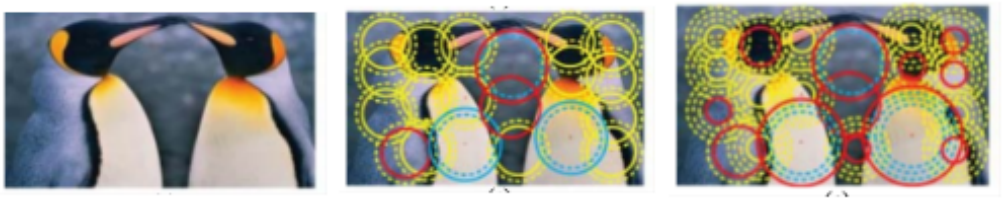


FIGURE 3.1: Visualization of Dynamic nCR by (Wei, Dai, and Zuo, 2016).

and Zuo, 2016) have not utilized the size dynamically adjusted receptive field size for any other process. In most of the cases, it has been used for better image representation (Wei, 2016) and saliency detection (Wei and Zuo, 2015).

In this work the information contained in this dynamic receptive field size is used to find edges in the input image. The basic intuition behind this is an edge or boundary point or region will have a inhomogeneous surround where as a non boundary region has a relatively homogeneous region. Further in case of dynamically finding RF size for a GC if the surround is inhomogeneous then according to the algorithm the size of the receptive field will shrink down to a much smaller value where as for a homogeneous region the adjusted RF size will be much higher. Hence from above two lines it could be inferred that a boundary region or an edge point will have a small receptive field size associated with it where as a non boundary region point will have a much larger value of adjusted RF size associated with it. This very inference serves as the base for the edge detection method proposed here. In this way if the adjusted size of the receptive field for each pixel is stored as color intensity value for this pixel in a new image. The edge or boundary region points will appears as black due to a smaller value and a non edge or non boundary point will be appearing as more white as following from the inference made above.

Hence, this inference can be modeled as an edge or boundary point detector. It is to be further noted that the particular value of adjusted receptive filed size not only indicate the nature of the pixel but also indicate about how big of a blob of the homogeneous region can be found around this pixel. This extended information can further be used for a blob detection system. The algorithm for edge or boundary point detection is given as follows:

Algorithm 2 Algorithm for edge detection using dynamic nCRF

1: Set image for edge detection as all black pixel value image of same size as that of input image
2: Set initial parameters min, max scale initial size for RF
3: **for** each central pixel for GC as pixel_value **do**
4: $size \leftarrow init_size$
5: $GC\_val \leftarrow image$ convolved with nCRF filter of size
6: Calculate GC_val-GC_val_small and GC_val-GC_val_large
7: **if** GC_val-GC_val_small differs from GC_val-GC_val_large **then**
8: **for** $size = init\_size - 1 : min\_size$ **do**
9: $GC\_val \leftarrow GC\_val\_size$
10: $GC\_val\_small \leftarrow GC\_val\_size - 1$
11: **if** GC_val not differs much from GC_val_small **then** break
12: **else**
13: **for** $size = init\_size + 1 : max\_size$ **do**
14: $GC\_val \leftarrow GC\_val\_size$
15: $GC\_val\_large \leftarrow GC\_val\_size + 1$
16: **if** GC_val differs much from GC_val_small **then** break
17: $intensity\_value\_this\_pixel \leftarrow GC\_val$
18: $intensity\_value\_this\_pixel\_new\_image \leftarrow size$

As explained earlier the algorithm 2 is mostly similar to algorithm1. The introduced concept for edge detection is done in Line 18 where the adjusted size of the filter is kept as intensity value for pixel position in edge map image.

3.4 Results

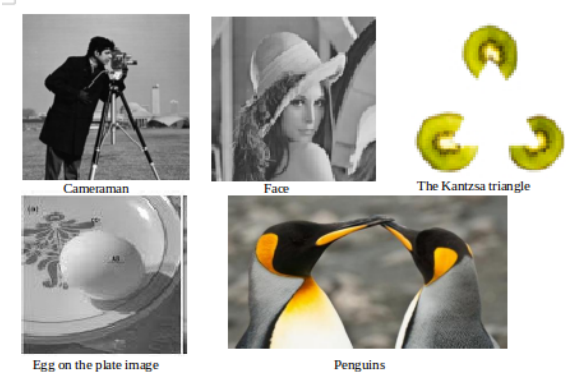


FIGURE 3.2: Various sample images.

(Wei, Wang, and Lai, 2012) have used the Y valued image of the Y, C_bC_r representation of the original image as a measure of comparison with the results they have got. In this work instead it is tried

in more visual manner. They further show some images to be subjected to segmentation which gave better results according to them. Here also, some filtered images are again subjected to segmentation and the results are inspected visually too. In the next two sections results for both the basic algorithm and edge detection version are discussed.Figure 3.2 shows some of the original images which have been used for the task. It contains images from simple and widely known domain like illusions as well as some segmentation benchmark image as egg on the plate.

3.4.1 Segmentation

This section discusses the effect of the dynamic nCRF method on image and its segmentation. The comparison between the adaptively filtering of original image, segmented result of original image using k means clustering and segmented result of the filtered image using the same method and values is presented in Figure 3.3.Some of the key observations are: We can clearly see the difference between segmented images for 1st row and 4th row images. In all the cases (specially in 1st and 4th row) the filtered image when segmented we can distinguish between object and background well even with similar colored object and back ground which is not there in other segmented image. The other observation is it can detect the region and counters shade effects too as can be clearly seen in 3rd and 5th images where despite shade regions are brought out in a better way. On the basis of these observations, we can say that a better segmentation result can be found through dynamic nCRF process.

3.4.2 Edge detection

Edge detection as discussed in section 3.3.2 is performed by the modified dynamic nCRF method. The result of sample images from Figure 3.2 subjected to the algorithm 2 is shown in Figure 3.4 along with the original and filtered images. The third column in Figure 3.4 shows the edge map for original images in 1st row. The expected edged could be clearly seen in the images. It could be observed that noise in the images are not considered as edges in all the figures.

Both the algorithms discussed above are expensive with respect to time because of its iterative nature. Along with size of image the running time also depends a lot on the threshold and initial values chosen for the algorithm. A small threshold means higher homogeneity is demanded within a region and vice-versa. A 240x150 image takes 5 minutes to get filtered along with a few more seconds to segment. For more heterogeneous images with smaller objects, the algorithm terminates quickly whereas more homogeneous images take more time to filter.

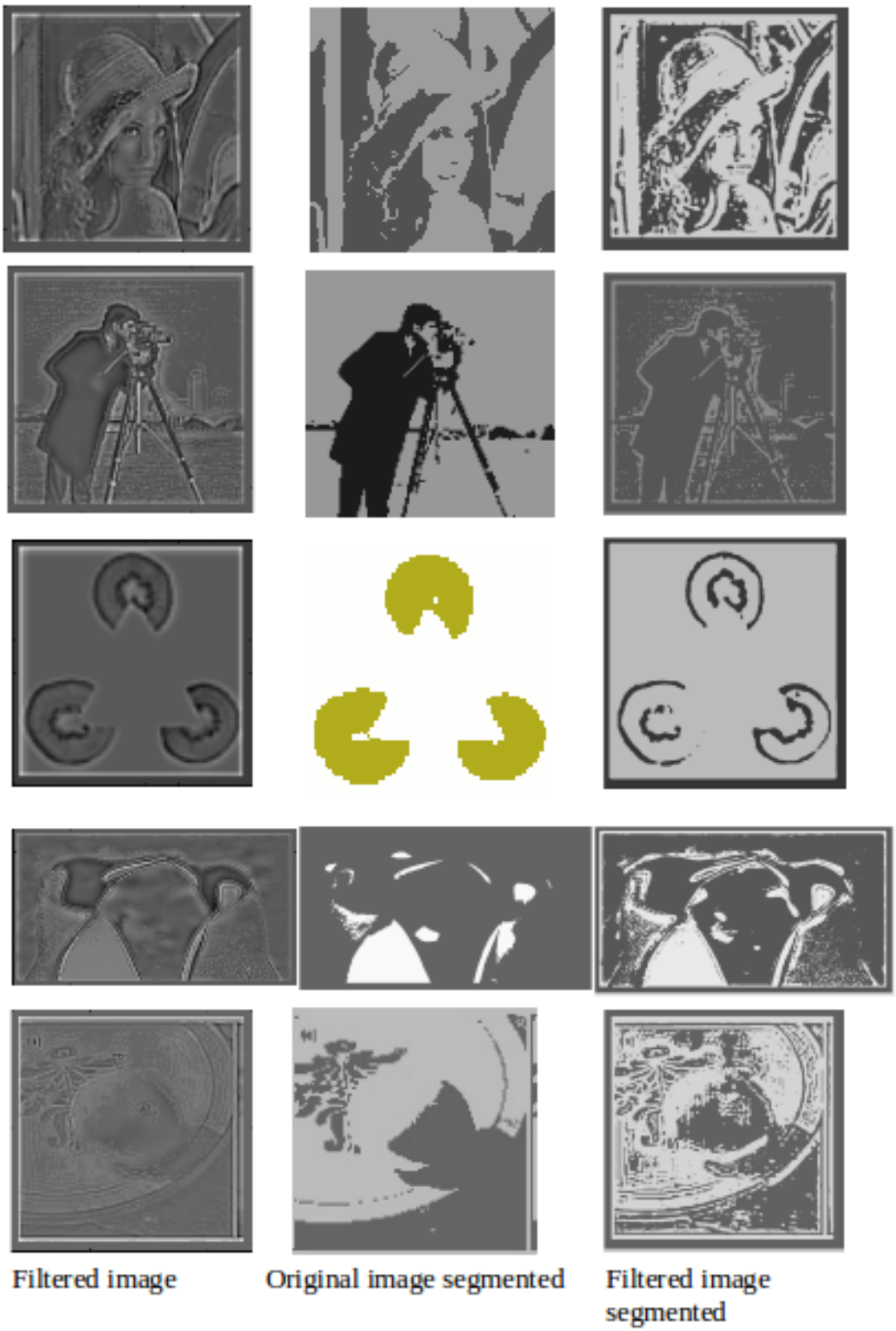


FIGURE 3.3: Segmented result comparison with Dynamic nCRF and without Dynamic nCRF

FIGURE 3.4: Edge detection result using Dynamic nCRF

Chapter 4

Impact of Convolutional Neural Network input parameters on classification performance

4.1 Abstract

Higher levels in visual hierarchy require more sophisticated models for performing higher level tasks like object recognition, face recognition and other classification tasks. A model for these tasks is usually complex and somewhat more abstract. All the models that has been discussed so far are shown to be biologically plausible ones and more inclined towards the mid level image representation. These are however, not very useful for higher level tasks like feature detection, object recognition, etc. From the pool of models for the higher level vision, deep Convolutional Neural Networks have shown impressive capabilities for solving complex image classification problems very well on a diverse range of problems in recent past. In the case of CNN there are numerous input parameters that decide the architecture of the network such as the number of convolutional layers, convolution kernel size, number of convolution filters in one layer, type of activation function, pooling window size, stride, etc. In this work an attempt is made to understand the impact of some of the input parameters on the classification performance of the network. The work is performed for a five class problem using a widely used color fundus retinal image dataset to classify stages of diabetic retinopathy. CNN input parameters such as the number of convolutional layers, number of filters in one layer, size of the convolution kernel and activation function is considered. The impact of these input parameters on the accuracy of classification and the runtime for training the network is analyzed. It has been observed that both the classification accuracy and the runtime

for training the network is more heavily dependent on the number of convolution filters in one layer and size of the convolution kernels than on the number of convolutional layers or the depth of the network. It is also found the type of activation function is actually having no impact on the accuracy. This preliminary work helps to understand the functioning of CNN, identify the crucial parameters which will finally lead to explanation of the reason behind their impact on the performance.

4.2 Introduction

Chapter 1 reviews a number of bio inspired models for different levels of vision. Chapter 2 and Chapter 3 uses the extensions or modification of the models for exploring aspects of how geometry affects size estimation via the Muller Lyer illusion and also understanding mid level vision to achieve goals like better mid level representation of image. All the models that has been discussed so far are shown to be biologically plausible ones and more inclined towards the mid level image representation. These are however, not very useful for higher level tasks like feature detection, object recognition, etc. Traditionally a number of attempts have been made in order to find bio motivated models which can perform well at these higher order tasks too. Some of these are Cellular neural network introduced by (Chua and Yang, 1988), spiking neural network by (Maass, 1997), Pulse coupled neural networks by (Johnson and Ritter, 1993) and others. Cellular neural nets are introduced as a parallel computing paradigm similar to neural networks, with the difference that communication is allowed between neighboring units only. Spiking neural nets on other hand make an effort to model the actions and firing phenomenon of neurons more accurately. Pulse coupled neural networks are neural models proposed by modeling a cat's visual cortex. Most of the models have been used in a variety of image processing applications, including: image segmentation, feature generation, face extraction, motion detection, region growing, and noise reduction. A similar bio inspired modeling of the visual system known as Convolutional Neural Networks(CNN) is one of the most studied topics in the community in recent past as shown by (Gu et al., 2017). It has been found to be very effective

in a number of domains and variety of problems. Its biological inspiration, ability to bring out relevant features from data and little requirement with respect to preprocessing of input data makes it a key player in the field. It needs to be noted also that the Receptive field modeling through spatial filters like DoG (low-level) or EDoG (mid-level), described in the earlier chapters, are also convolutional networks in a primitive form. A common issue for all of these models, mentioned here, is the choice of the non trainable parameters of the model. In the study of each of the models attempts have been made to understand the nature of these non trainable parameters. Similar attempts have also been made in the field of Convolutional neural nets and deep learning too. In this work also an attempt has been made to understand the nature of these non trainable parameters with accuracy and training time for a classification problem. A good understanding of the relevance of these will definitely help the community to understand the convolutional neural networks better. It would help to understand logically what parameter value to be chosen for a given problem. To execute this job, an extensive study has to be done on the various possible values of these non trainable parameters on a certain dataset in a set up for classification problem. In the next sections the various constituents of a convolutional neural net as well as the terminology for the non trainable parameters of a model under consideration here is discussed along with the detail about the task on which the investigation is to be done.

4.2.1 Types of layers

A brief introduction about each type of layer and how they have been used in this work is given as follows:

Convolutional Layer

This layer works very similar to the prototype of a layer in the sense that it has learn able weights and biases. It receives input, performs a dot product and allows an optional non-linearity. The main difference is that it takes a locality into account while calculating activation for a particular position in the input using the concept of the

(A) Convolution layer. (B) Max pooling layer.

FIGURE 4.1: Convolution and max pooling layer visualization.

receptive field discussed in Section 1.2.2. The inner structure consists of a set of filters(kernel) which means the neurons are arranged in form of 3D structures to applicable on 3D image shown in Figure 4.1a. The convolution of the input image with each filter produces features which are extracted by forming a new layer. Each layer transforms the 3D input volume to a 3D output volume. Each layer represents projection of input data into some higher dimensional space.

Max-pooling Layer

Max-pooling is discretization process. It is used to down-sample an input representation resulting in reducing it's dimensionality and reduced number of parameters to learn. Further reducing the computational cost. The features contained in a sub region could be abstracted by translational invariance property. This layer is generally used after a convolutional layer. The action of a 2X2 max-pooling is shown in Figure 4.1b.

Activation Layer

In order to add nonlinearity after each layer this layer is used. Without nonlinearity the whole network act as a simple linear transformation. It is known that the linear networks do not have so much power for the complicated task such as image classification. A number of nonlinear activation function have been prevailing. We have used some of them for our analysis. Sigmoid, Tanh and ReLU are the three activations functions used here for investigation. Each of these has a strong biological correspondence. The functions are

given as follows:

$$Sigmoid \qquad f(x) = \frac{1}{1+e^{-x}}$$

$$Tanh \qquad f(x) = Tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

$$ReLU \qquad f(x) = max(0,x)$$

Dropout Layer

Dropout is a regularization technique for reducing over fitting in neural networks by preventing too much reliance on training data (Srivastava et al., 2014). Here, while updating a neural net layer, it updates each node with probability p, and leave it unchanged with probability 1-p. Hence, only the reduced network is trained on the data in that stage. The removed nodes are then reinserted into the network with their original weights.

Fully Connected Layer

The fully connected layers are used to perform the higher level reasoning on the features extracted by use of stack of convolutional and max pooling layers. This layer takes all input from each neuron from the previous layer and connects to each and every neuron of the next layer.

Classification Layer

After a stack of multiple layers, the final layer is a classifier layer which is stacked on the top for classifying the input. Softmax is a popular choice for the classification function. This function has been used for classification amongst classes in this work also.

$$L_i = -\log(\frac{e^{f_i}}{\sum_j e^{f_j}})$$

where f j is the j-th element of the vector of class scores f . A correct prediction probability is achieved by the softmax in the log of the equation.

4.2.2 Hyper parameter and related works

Traditionally the term **hyperparameter** refers to a parameter whose value is set before the learning process begins in machine learning.In contrast the values of other parameters are derived via training. In this work intension is to find out the nature of these hyper-parameters. Hence, these are also treated as the parameters to be found via analysis in this work.

In recent times a seminal work in neural networks was in 2012 when (Krizhevsky, Sutskever, and Hinton, 2012) built a deep CNN model called AlexNet. It consists of five convolutional layers and three fully connected layers to train the ImageNet dataset for a image recognition challenge. It obtained state-of-the-art performance on the ImageNet dataset. It followed with a number of other major architectures like VGGNet by (Simonyan and Zisserman, 2014), ResNet by (He et al., 2016) and many more. The hyper-parameters in all of these work are chosen in a more of trial and error method. More recently approaches have been made to understand the relevance of the parameters with accuracy and time required for training. Albelwi et al. used objective function that combines the information from the visualization of learned feature maps to provide a framework for choosing hyper-parameter of the network by (Albelwi and Mahmood, 2016). Sequential model based optimization (SMBO) has been used for the same job for object recognition by (Talathi, 2015).(Kim, 2014) used convolutional neural networks (CNN) trained on top of pre-trained word vectors for sentence-level classification tasks. It also examines the relevance of hyper-parameters with accuracy for the problem. Extrapolation of learning curves has been used by (Domhan, Springenberg, and Hutter, 2015) to automatic hyper-parameter optimization. (Basu et al., 2015) proposed a framework for the satellite imagery named DeepSat. They have used a number of techniques on a number of satellite images datasets to produce the learning framework.

4.2.3 Diabetic Retinopathy

Problems related to medical imaging in this regard are of a significant interest. Since medical image analysis has seen notable interest in recent past from community it has been chosen as the area for

analysis here. It is assumed that if the analysis is done on a certain specialized type of problem and dataset then more strong inference can be made out of it.

Diabetic Retinopathy(DR) is a medical condition in which due to prolonged diabetic mellitus some defects in retinal vision is observed. This damaged retina can lead to blindness. It affects up to 80% of people having diabetes for more than 20 years. It is estimated that diabetes mellitus affects 4 per cent of the world's population, almost half of whom have some degree of DR at any given time as in (Kertes and Johnson, 2007). India has highest number of diabetes mellitus with 31.7 million people, China is second with 20.8 million people and USA is third with 17.7 million people as shown in (Kaveeshwar and Cornwall, 2014). The detection of DR requires the examination of the color fundus photographs of retina by trained clinicians which is a time consuming work. Further, the clinicians require specialized fundus cameras to capture the photograph of the retina.

Different machine learning techniques have been applied to classify the different stages of DR using the color fundus images. The automated techniques for identification of the stages of DR have been developed by support vector machines in (Giraddi, Pujari, and Seeri, 2015) and k-NN classifiers in (Mookiah et al., 2013). Recursive region growing segmentation algorithms in (Sinthanayothin et al., 2002) and artificial neural network has also been done in (Usher et al., 2004). A random forest based model has been used for the same problem by (Casanova et al., 2014). Most of the models work by involving preprocessing to standardize color and enhance contrast. The classification accuracy provided by most of the models are encouragingly high for the given datasets. A similar but slightly mixed approach has been taken by top-down image segmentation and local thresholding by a combination of edge detection and region growing in (Jaafar, Nandi, and Al-Nuaimy, 2011). More recent work provided a very significant result with the use of the deep neural network and CNN by (Ghosh, Ghosh, and Maitra, 2017). It uses 27 layers deep neural network to provide a high accuracy on both 2 class and 5 class variants of the problem.

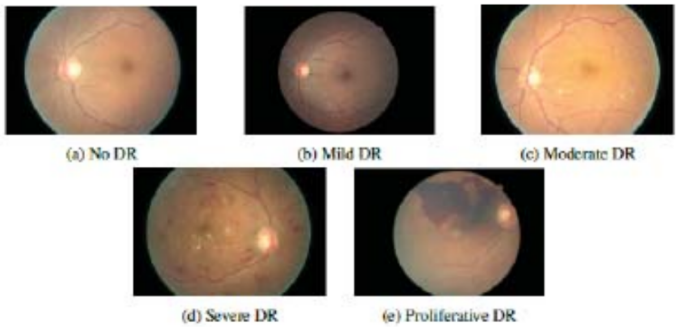


FIGURE 4.2: Sample images from DR dataset.

4.2.4 Dataset

Data is used from the dataset made available by the Kaggle competition website and maintained by EyePacs (*DR dataset website Kaggle*). The dataset consists of color fundus photographs of the retina. The severity of DR is the basis for the classification. In the dataset a trained clinician assigned each image to a class. The class labels of the dataset are highly imbalanced i.e. more than 73.5% of the class are negative. The class partition of the original 35000 images provided for training purpose is given in Table 4.1, where PDR and NPDR refers to proliferative and Non-proliferative DR respectively. The classification task was to classify a given fundus image as to one of the five classes. Some of the images from each class are shown in Figure 4.2.

Class	Instances	Percentage
Negative	25810	73.5%
Mild NPDR	2443	6.9%
Moderate NPDR	5292	15.1%
Severe NPDR	873	2.5%
PDR	708	2%

TABLE 4.1: Class partition for DR dataset

4.3 Methodology

The main task in this work is to investigate the relevance of the hyper parameters with respect to accuracy and time consumed for training one epoch of data. Since an investigation involving all possibilities of the hyper-parameter is not feasible computationally, we therefore settle here for a subset showing prominent behavior of

variation. The main hyper-parameters which are dealt with here are the depth of networks number of convolutional layers in particular, number of filters or neurons in the network, size of filters in convolution layers and activation function.

4.3.1 Preprocessing

The size of images in the dataset is unequal throughout and rather high for efficient computation. Hence, it is reduced to size 512X512 via cropping and resizing. The cropping is done manually in such manner to keep only retinal image and then subjected to automate resizing. Non-Local Means De-noising (NLMD) is implemented as the prepossessing step introduced by (Buades, Coll, and Morel, 2005). The de-noising of an image x =(x1; x2; x3) at pixel j on channel i is given as:

$$X_i(j) = \frac{\sum_{k \in B(j,r)} X(j).W(j,k)}{C(j)}$$

$$C(j) = \sum_{k \in B(j,r)} W(j,k)$$

where B(j; r) is a neighborhood around pixel j with radius r, and the weight w(j; k) is the square of Frobenius norm distance between color patches centered at j and k that decays under a Gaussian kernel.

4.3.2 Optimization function

Adam stands for adaptive moment estimation introduced by (Kingma and Ba, 2014). It is an adaptive optimization algorithm to update network weights iterative based in training data. The procedure maintains a separate learning rate for each network weight and adapts as learning unfolds. In this way it contains the advantages of both AdaGrad and RMSProp algorithms. It uses both first and second moments of the gradients. In this work a learning rate of 1e-4 has been used. For other parameters like exponential decay rates and epsilon default values are used.

4.3.3 System specification and Evaluation metric

Dell Precision Tower T7810 is used for computation having Intel Xeon 2.4 GHZ processor with 128GB RAM and 12GB Nvidia Titan X GPU. The widely used open-source library TensorFlow is used for implementing the CNN. Out of the complete dataset of 35000, a portion of 80% has been used for training purpose and remaining 7000 has been used for testing purpose. Since we are computing for only one epoch no validation data has been kept. All the three channels of the original image have been kept intact. Accuracy in our case has been defined as the proportion of the samples correctly classified i.e.

$$Accuracy = \frac{\text{Number of correctly classified samples}}{\text{Number of total samples}}$$

Time required in our case is defined as the time spent for training one epoch of the dataset on the hardware with no other major process running simultaneously on it.

4.3.4 Conventions and Implementation details

The filter as well as the fully connected edge weights in all layers in each of the architecture tested are all initialized randomly. It is done so to avoid any effect of special initialization process since we are not examining for the initialization method in this work.

The convention for representing an architecture configuration for a layer with k convolutional layers and n fully connected layers is as follows:

X1-Y1:X2-Y2:X3-Y3: :Xk-Yk::A1:A2: :Ak

Where Xi=number of filter in layer I
Yi=size of filter in layer i
Ai= number of neuron in fully connected layer I
: marks end of a layer and start of another
:: marks end of convolutional layer and start of fully connected layers

For the purpose of feasible computation complexity only a particular configuration of the fully connected layer containing 2 fully

connected layer each with 128 units or neurons have been used. As stated earlier main focus is on convolutional layers. Further one of the two fully connected layer is connected to one layer of dropout. Out of the two fully connected layer used here the one next to classification layer is chosen for dropout. The dropout is applied with a keep probability p=0.25.

4.4 Results

The results and inference from the analysis of the subjected hyper-parameter are presented next parameter-wise.

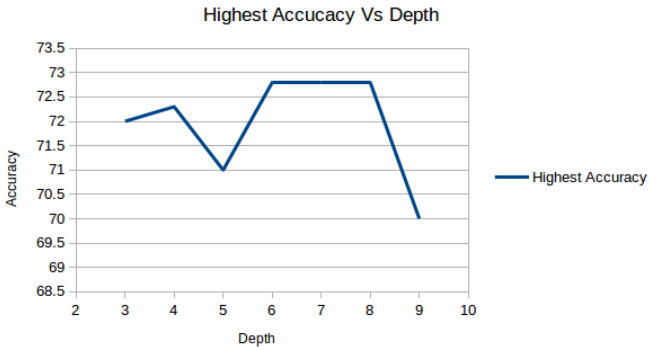


FIGURE 4.3: Plot for Highest Accuracy Vs Depth.

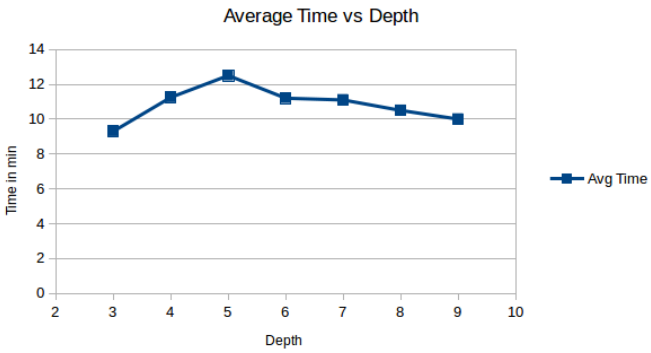


FIGURE 4.4: Plot for Average time Vs Depth.

4.4.1 Depth or Number of convolution layers

The effect of depth on highest accuracy achieved for the problem is shown as a line graph in Figure 4.3. The trends show that a similar level of accuracy could be achieved with different levels. Further, it

remains high for a moderate depth and rather less for a low or high number of depths.

The effect of depth on training time for one epoch for the problem is shown as a line graph in Figure 4.4. The trends show that with increasing depth the time consumed for training increases.

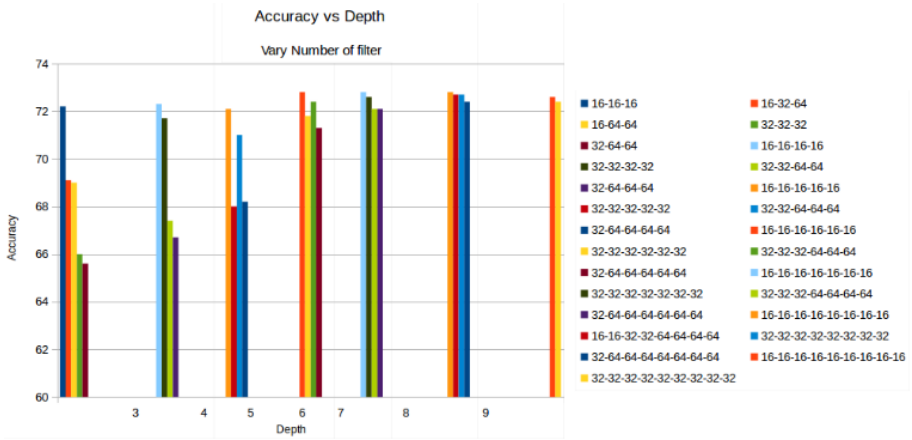


FIGURE 4.5: Bar graph for Accuracy Vs Depth for varying number of filters.

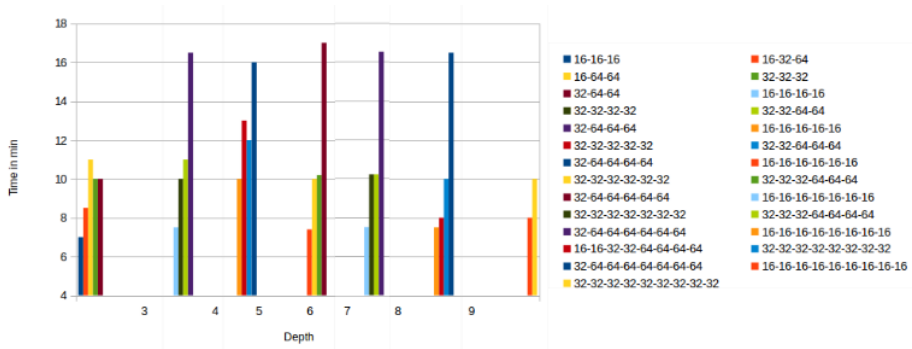


FIGURE 4.6: Bar graph for Time Vs Depth for varying number of filters.

4.4.2 Number of filters

The effect of variation of number of filters with accuracy is shown in Figure 4.5 for a number of depths and configurations. In the given bar graph each colored bar represents one architecture. The plot shows that with increasing number of filters the accuracy decreases. This could be because too many filters might over-fit the data.

The effect of variation of number of filters with time is shown in Figure 4.6 for a number of depth and configuration. Similar to Figure 4.5 in the given bar graphs each colored bar represents one architecture. The plot shows that with increasing number of filters

the time increases. This could be understood because larger number of filters mean increase in number of parameters in the network. This will lead to increased time for training.

4.4.3 Size of filters

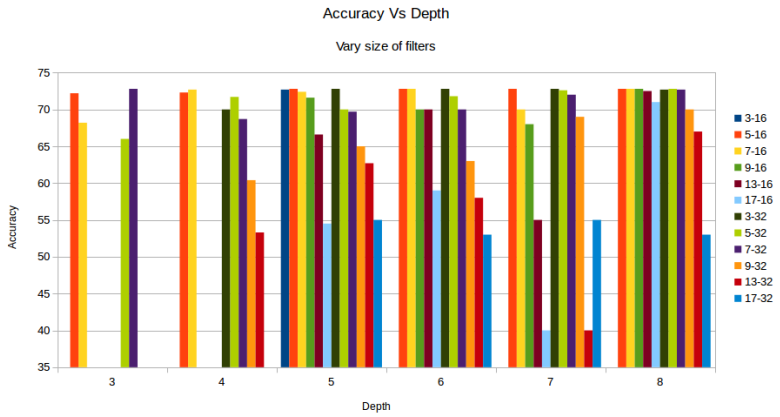


FIGURE 4.7: Bar graph for Accuracy Vs Depth for varying size of filters.

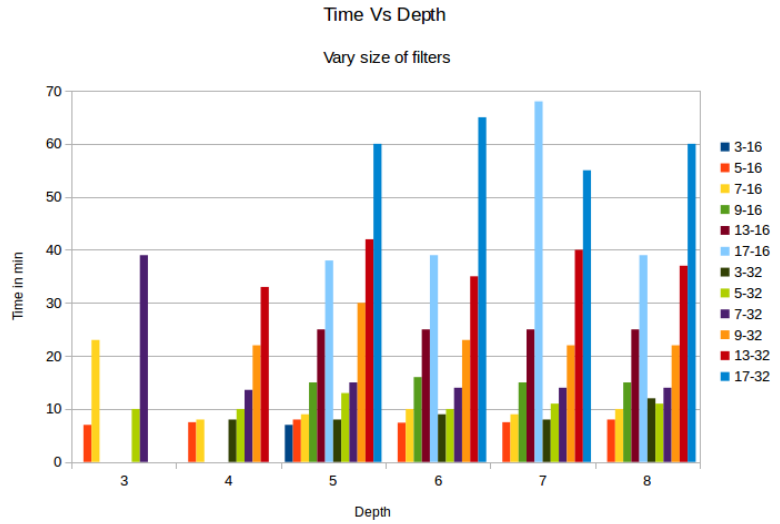


FIGURE 4.8: Bar graph for Time Vs Depth for varying size of filters.

The effect of variation of size of filters with accuracy with respect to the number of layers is shown in Figure 4.7. Each of the colored bars is represented in the form x-y indicating all the layers have filter of size x and each convolution layer is having y number of filters. For each depth, filter size 3,5,7,9,13 and 17 are tested. 16 and 32 filters are used. The plot shows that with increasing size of the filter the accuracy decreases for most of the depths.

The effect of variation of size of filters with accuracy is shown in Figure 4.8 for a number of depths. Colored bars hold same mining as earlier. The plot shows that with increasing size of the filter the training time increases.

4.4.4 Activation functions

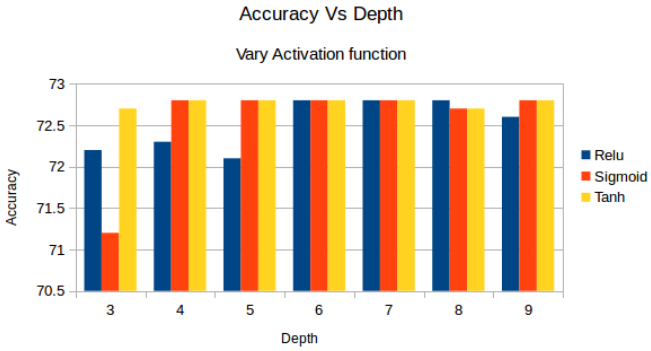


FIGURE 4.9: Bar graph for Accuracy Vs Depth for varying Activation functions.

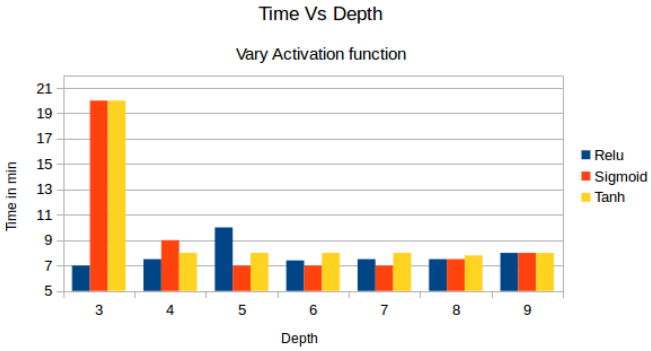


FIGURE 4.10: Bar graph for Time Vs Depth for varying activation functions.

The effect of variation of activation function with accuracy is shown in Figure 4.9 for a number of depths. The 3 colored bars indicate three activation functions used here. For each depth a configuration in which all layers have filter size of five and number of filters 16 has been used. The plot shows no real significant change in accuracy with respect to activation functions.

The effect of variation of activation function of filters with training time is shown in Figure 4.10 for a number of depths. The significance of bars and configuration are same as earlier plot. The plot

shows no real significant change in training time with respect to activation functions except for the three convolution layered network where it is unexpectedly high for sigmoid and Tanh functions.

This preliminary work, done here, helps to understand the functioning of a deep CNN, and identify the crucial parameters which will finally lead to explanation of the reason behind their impact on the performance. This, in turn, may lead to a better understanding of the organization of visual structures and their functions, how the three-dimensionsl world view emerges from a two-dimensionsional intensity array on the retina, which pathways and connections are crucial, and which are not and so on. For instance, the LGN was originally envisaged as a relay station between eye and cortex that transmits the weakend signals amplified, but now it has emerged as a crucial point of interaction between bottom-up and top-down signals; in fact, very surprisingly, the number of cortico-geniculate synapses far exceed the retino-geniculate synapses in LGN. Understanding the process of deep learning may amount to taking some firm steps in realizing the processes of perception and cognition in the brain.

Chapter 5

Conclusion and Future work

A summery of the models studied in this work its correspondence with Marr's hierarchy and their applications is given in Table5.1.

Maar's Hierarchy	Bio motivated model	Purpose
Primal Sketch	DoG	Edge detection
2.5D sketch	EDoG	Segmentation; size and edge estimation
3D	deep CNN	Classification

TABLE 5.1: Correspondence of models used and Marr's hierarchy

The main idea of the thesis work was to understand the hierarchical approach to vision by (Marr, 1982) and present corresponding biologically motivated models. In course of the study, the existing models have been used for different application. Modification in the existing models have been made to get more features from an image as well. The central idea of the thesis was to continue the quest for a unified model of vision. Although it has, in a sense, been achieved, a number of problems still awaits to be examined by the models discussed. The future work for this is to find similar structures which can account for a unified model for visual. The conclusion made from individual chapter work has been given, along with future works, for clarity. Nonetheless, the direction of future work remains the same, viz an approach towards more biologically plausible and useful networks.

In Chapter 2 an attempt has been made to explain the Muller Lyer illusion with the help of nCRF model and contour plots. Further The relation amongst the arrows angles and the induced illusion is explored. The results found by computational methods proposed are compared with the psychophysical experimental data. The results obtained from the computational model proposed to explain the original Muller Lyer illusion are promising. The computational method proposed finds the perceived length of both the lines and their difference close to the experimental data. Further the attempt

to find the relation between arrow angle and induced illumination are also encouraging. The curve for the relation from both the proposed approach and the experimental data looks close to each other. Hence, it could be conclusively said that the proposed nCRF based model is a good tool to understand how size is affected from shape and geometry of and around an object. The future work includes exploring other aspects of space cognition with the help of the proposed method. It includes how other parameters of the MLI, like length of the line or wings, relative brightness of the line colors, etc affects the induced illusion. The application of the nCRF model to explain similar geometrical illusions has also been planned. This

In Chapter 3 the concept of dynamic ECRF, a modeling technique for incorporating the top down and bottom up approaches in vision have been discussed. In particular the work by (Wei, Wang, and Lai, 2012)has been understood and implemented for getting a better mid level representation of an image through the use of the dynamic ECRF algorithm. Further, an edge detection algorithm has been proposed by using the size of adaptive receptive field information provided by the algorithm by (Wei, Wang, and Lai, 2012). This results in providing a model which can simultaneously give a mid-level representation along with edge map of an image. This feature can prove to be very useful for jobs like motion detection, where outline detection and object detection is required simultaneously.

Encouraging results have been observed for both segmentation as well as edge detection algorithms visually in Section 3.4. In spite of the model's ability to provide better visual results, time consumption, because of the iterative nature of the algorithm, is a big concern. The future work in this regard is to modify the algorithm to decrease time consumption. Another aspect includes verifying the result and performance on some hand labeled datasets. The adaptive nature of the algorithm is a nice concept, it is planned to use this concept with other image filtering methods.

As an approach to higher level of vision in Marr's hierarchy in chapter 4 the basic components of the convolutional neural network and deep learning is discussed. An investigation onto the relevance of various hyperparameters with the accuracy and training time for a particular dataset has been performed. The results show the relation of the hyperparameters with the accuracy as well as training

time for the classification problem. Future work includes doing this extensive study on a more generalized dataset and produce more inferences from the work. Another aspect for the future work is to use bio inspired initialization for the filters of CNN. The filters could be initialized with Gaussian, DoG or a combination of other such filters for the purpose.

Bibliography

Albelwi, Saleh and Ausif Mahmood (2016). "Automated optimal architecture of deep convolutional neural networks for image recognition". In: *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. IEEE, pp. 53–60.

Basu, Saikat et al. (2015). "Deepsat: a learning framework for satellite imagery". In: *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, p. 37.

Blakeslee, Barbara and Mark E McCourt (1999). "A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction". In: *Vision research* 39.26, pp. 4361–4377.

Buades, Antoni, Bartomeu Coll, and J-M Morel (2005). "A non-local algorithm for image denoising". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 2. IEEE, pp. 60–65.

Bulatov, Aleksandr, Algis Bertulis, and L Mickienė (1997). "Geometrical illusions: study and modelling". In: *Biological Cybernetics* 77.6, pp. 395–406.

Casanova, Ramon et al. (2014). "Application of random forests methods to diabetic retinopathy classification analyses". In: *PLOS one* 9.6, e98587.

Chua, Leon O and Lin Yang (1988). "Cellular neural networks: Applications". In: *IEEE Transactions on circuits and systems* 35.10, pp. 1273–1290.

Curcio, Christine A et al. (1990). "Human photoreceptor topography". In: *Journal of comparative neurology* 292.4, pp. 497–523.

Daugman, John G (1985). "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters". In: *JOSA A* 2.7, pp. 1160–1169.

Davson, Hugh and Edward S. Perkins (2018). *Human eye*. URL: https://www.britannica.com/science/human-eye.

Day, Ross H and Hannelore Knuth (1981). "The Contributions of F C Müller-Lyer". In: *Perception* 10.2. PMID: 7024901, pp. 126–146. DOI: 10.1068/p100126. eprint: https://doi.org/10.1068/p100126. URL: https://doi.org/10.1068/p100126.

Domhan, Tobias, Jost Tobias Springenberg, and Frank Hutter (2015). "Speeding Up Automatic Hyperparameter Optimization of Deep Neural Networks by Extrapolation of Learning Curves." In: *IJCAI*. Vol. 15, pp. 3460–8.

DR dataset website Kaggle. https://www.kaggle.com/c/diabetic-retinopathy-detection. Accessed: 2018-06-22.

Ghosh, K, S Sarkar, and K Bhaumik (2005a). "Low-level brightness-contrast illusions and non classical receptive field of mammalian retina". In: *Intelligent sensing and information processing, 2005. Proceedings of 2005 international conference on*. IEEE, pp. 529–534.

Ghosh, Kuntal, Sandip Sarkar, and Kamales Bhaumik (2005b). "A possible mechanism of zero-crossing detection using the concept of the extended classical receptive field of retinal ganglion cells". In: *Biological Cybernetics* 93.1, pp. 1–5.

— (2006). "A possible explanation of the low-level brightness–contrast illusions in the light of an extended classical receptive field model of retinal ganglion cells". In: *Biological Cybernetics* 94.2, pp. 89–96.

Ghosh, Ratul, Kuntal Ghosh, and Sanjit Maitra (2017). "Automatic detection and classification of diabetic retinopathy stages using CNN". In: *Signal Processing and Integrated Networks (SPIN), 2017 4th International Conference on*. IEEE, pp. 550–554.

Giraddi, Shantala, Jagadeesh Pujari, and Shivanand Seeri (2015). "Identifying abnormalities in the retinal images using svm classifiers". In: *International Journal of Computer Applications* 111.6.

Gu, Jiuxiang et al. (2017). "Recent advances in convolutional neural networks". In: *Pattern Recognition*.

Hartline, H K, Henry G Wagner, and Floyd Ratliff (1956). "Inhibition in the eye of Limulus". In: *The Journal of general physiology* 39.5, pp. 651–673.

He, Kaiming et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hubel, David H and Torsten N Wiesel (1959). "Receptive fields of single neurones in the cat's striate cortex". In: *The Journal of physiology* 148.3, pp. 574–591.

— (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1, pp. 106–154.

Jaafar, Hussain F, Asoke K Nandi, and Waleed Al-Nuaimy (2011). "Automated detection and grading of hard exudates from retinal fundus images". In: *Signal Processing Conference, 2011 19th European*. IEEE, pp. 66–70.

Johnson, John L and Dieter Ritter (1993). "Observation of periodic waves in a pulse-coupled neural network". In: *Optics letters* 18.15, pp. 1253–1255.

Kaveeshwar, Seema Abhijeet and Jon Cornwall (2014). "The current state of diabetes mellitus in India". In: *The Australasian medical journal* 7.1, p. 45.

Kawabata, Nobuo (1976). "Mathematical analysis of the visual illusion". In: *IEEE Transactions on systems, man, and cybernetics* 12, pp. 818–824.

Kertes, Peter J and Thomas Mark Johnson (2007). *Evidence-based eye care*. Lippincott Williams & Wilkins.

Kim, Yoon (2014). "Convolutional neural networks for sentence classification". In: *arXiv preprint arXiv:1408.5882*.

Kingma, Diederik P and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105.

Kuffler, Stephen W (1953). "Discharge patterns and functional organization of mammalian retina". In: *Journal of neurophysiology* 16.1, pp. 37–68.

Li, CY (1997). "New advances in neuronal mechanisms of image information processing". In: *Bulletin of National Natural Science Foundation of China* 3, pp. 201–204.

Maass, Wolfgang (1997). "Networks of spiking neurons: the third generation of neural network models". In: *Neural networks* 10.9, pp. 1659–1671.

Mandal, Mainak (2016). "Geometric Optical Illusions". MA thesis. Indian Institute of Science Education and Research, Kolkata: Depatrment of Mathematics.

Marr, David (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY, USA: Henry Holt and Co., Inc. ISBN: 0716715678.

Marr, David and Ellen Hildreth (1980). "Theory of edge detection". In: *Proc. R. Soc. Lond. B* 207.1167, pp. 187–217.

McIlwain, James T (1966). "Some evidence concerning the physiological basis of the periphery effect in the cat's retina". In: *Experimental Brain Research* 1.3, pp. 265–271.

Mookiah, Muthu Rama Krishnan et al. (2013). "Computer-aided diagnosis of diabetic retinopathy: A review". In: *Computers in biology and medicine* 43.12, pp. 2136–2155.

Restle, Frank and James Decker (1977). "Size of the Mueller-Lyer illusion as a function of its dimensions: Theory and data". In: *Perception & Psychophysics* 21.6, pp. 489–503.

Simonyan, Karen and Andrew Zisserman (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*.

Sinthanayothin, Chanjira et al. (2002). "Automated detection of diabetic retinopathy on digital fundus images". In: *Diabetic medicine* 19.2, pp. 105–112.

Srivastava, Nitish et al. (2014). "Dropout: A simple way to prevent neural networks from overfitting". In: *The Journal of Machine Learning Research* 15.1, pp. 1929–1958.

Talathi, Sachin S (2015). "Hyper-parameter optimization of deep convolutional networks for object recognition". In: *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, pp. 3982–3986.

Usher, Dumskyj et al. (2004). "Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening". In: *Diabetic Medicine* 21.1, pp. 84–90.

Webb, Ben S et al. (2002). "Feedback from V1 and inhibition from beyond the classical receptive field modulates the responses of neurons in the primate lateral geniculate nucleus". In: *Visual neuroscience* 19.5, pp. 583–592.

Wei, Hui (2016). "A bio-inspired integration method for object semantic representation". In: *Journal of Artificial Intelligence and Soft Computing Research* 6.3, pp. 137–154.

Wei, Hui, Zhi-Long Dai, and Qing-Song Zuo (2016). "A ganglion-cell-based primary image representation method and its contribution to object recognition". In: *Connection Science* 28.4, pp. 311–331.

Wei, Hui, Xiao-Mei Wang, and Loi Lei Lai (2012). "Compact image representation model based on both nCRF and reverse control mechanisms". In: *IEEE transactions on neural networks and learning systems* 23.1, pp. 150–162.

Wei, Hui and Qingsong Zuo (2015). "A Biologically Inspired Neurocomputing Circuit for Image Representation". In: *Neurocomput.* 164.C, pp. 96–111. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2015.01.078. URL: http://dx.doi.org/10.1016/j.neucom.2015.01.078.