

Scale- and Translation-Invariant Unsupervised
Learning of Hidden Causes Using Spiking
Neurons with Top-Down Attention

Youssef Kashef

7. August 2013

Contents

Chapter 1

Abstract

Nessler et al. have demonstrated the ability of a spiking neuronal network governed by spike-timing-dependent-plasticity (STDP) and a stochastic winner-take-all (WTA) circuit to learn and predict causes from visual input. We aim to increase the computational power of the existing network through invariance to translation and scale. The visual system of the brain masters the recognition of objects wherever they appear in the visual scene and regardless of scale, orientation or even with partial occlusions. It achieves this through attention. Therefore, we turn to the pool of literature on modeling visual attention systems inspired from the brain. The architecture of the extended model is composed of the existing recognition module whose response modulates the attention module to be constructed in a top-down manner. This modulation will allow the attention module to alter the input window exposed for recognition. Attention is modeled as a network measuring for saliency in a scene by feature extraction with the use of hierarchies. The design and development of this extended model to achieve the required invariance using processes that approximate their biological counterparts is presented. Emphasis is put on making these approximations through computationally economic implementations. Evaluation of the model is based on its performance in a set of experiments as well as its computational efficiency. Experiments are constructed to scrutinize the behavior of the model, its ability to converge onto a sight within a scene that enables recognition. Artificial as well as natural images are used to further reveal the capabilities and limitations of our approach.

Chapter 2

Introduction

elaborated abstract with references.....[?]

Chapter 3

Object Recognition with Spike-based Expectation Maximization

3.1 Spike-based Expectation Maximization

Literature review of existing SEM model

Nessler et. al have articulated a model of bayesian modules of how the brain analyzes sensory stimuli. The model demonstrates the learning of hidden causes in visual stimuli emerging through correlations in a stochastic soft winner-take-all (WTA) network of spiking neurons activated continuously in the presence of their preferred stimulus [?]. It utilizes spike-timing-dependent-plasticity (STDP) in WTA circuits as an approximation of Expectation Maximization [?]. This model forms the basis of the presented work.

More emphasis will be put on how the WTA circuit in the SEM model, used as our main building block, is constructed. It comprises of a feed-forward single layer spiking neural network. The input layer is made out of spiking nodes whose firing activity is governed by a poisson process. External variables undergo a population coding that determine the modality of the poisson process. In the example offered by Nessler. The external variables are intensity values (pixels) form a static visual stimulus (image). The population coding polarizes these intensity values into binary on-off states which directly determines the firing probability of the poisson process. A spiking neuron is assigned to encode each state of the population code generate for each pixel. The firing rate of these neurons will be proportional to the state of the node in the population code.

3.2 SEM for learning features

3.2.1 Extending SEM by learning orientations

The current encoding of external variables accounts for the intensities of the spatial units (pixels) of a presented stimulus. The encoding of intensities is performed through a population coding by antagonistic binary nodes per pixel that drive a poisson process [?]. Parallel to these intensity encoded nodes, we add a WTA circuit per pixel that determines the preferred orientation of this node relative to its spatial neighbors. This creates an orientation map of the presented stimulus. Whilst counter-intuitive with traditional learning models, SEM benefits from elaborating the dimensionality of WTA s feature space as this increases its resolution for detecting correlations between an output node z and input nodes y on a linear scale. Although practically an increase in dimension, which the SEM aims to reduce, it is preferable to describe it as an elaboration of dimensionality since the added dimensions, or nodes, do not carry any new information, but rather refine its representation. Recalling the use of using population coding to encode in antagonistic (on-node, off-node) fashion, thus letting the WTA learn the likelihood of an input node firing, or not firing, explicitly, as shown by ??.

$$p(z = 1|y) \propto y * p(y = 1|z) + (1 - y) * p(y = 0|z) \quad (3.1)$$

where

- z denotes an output node,
- y denotes an input node

As we introduce the orientation map we may add additional operands to ?? to account for the nodes preferred orientation.

$$p(z = 1|y_I \cup y_O) \propto y_I * p(y_I = 1|z) + (1 - y_I) * p(y_I = 0|z) + y_O * p(y_O = O_p|z) + \sum_{i \neq O_i} (1 - y_O) p(y_O = o_i|z) \quad (3.2)$$

where

- y_I denotes an input intensity node,
- y_O denotes an orientation input node,
- O denotes the set of orientations available. Orientations can be defined discretely and arbitrarily (e.g. 30, 60,...180 degrees) or they can be learned [?],
- O_p denotes the preferred orientation

We redesign the network with a cascade of hierarchical WTA circuits. The input layer is a matrix of WTA circuits per spatial. Each input WTA circuit decides on the preferred orientation and intensity of its input. We will experiment with configuring the input WTA circuit to only relay intensity, only orientation, or both information.

The WTA circuits responsible to determine the preferred orientation from a set of predefined discrete set of orientations for each spatial node are activated by convolving the stimulus with a bank of two-dimensional Gabor filters. The filters are defined with different angular orientations and scales. By comparing the magnitude of responses between the filters at each pixel we can decide on its preferred orientation. It is worth mentioning that talking about the preferred orientation if a pixel would not make a lot of sense. The transformations do yield responses for each pixel but they only becomes informative in relation to the responses of its neighbors.

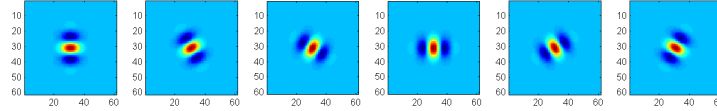


Figure 3.1: Example of Gabor filters (real part). Defined at orientations $[0, 150]$ degrees with 30 degrees increments.

Parameterized Gabor functions are an adequate approximation of simple cells in the primary visual cortex [?]. Although Daugman demonstrates the construction of a neural network to achieve this transformation this work adopts the traditional systems' approach for defining and applying the filters [?].

3.3 Extending SEM for learning hidden features

extending SEM by learning abstract features

We have seen the computational power of the SEM model as an unsupervised method for identifying hidden causes. So far the hidden causes have been used synonymously with predefined classes (e.g. digits[?]). We will extend the SEM model in a way that breaks this assumption. We insert an additional WTA circuit, responsible for learning hidden causes that depict abstract features of the object we're attempting to detect and recognize. This feature layer will contribute to the bottom-up learning as we expose it to the low-level input and have it drive the WTA circuit already encountered in the original SEM architecture. With this additional feature-WTA circuit introduced we no longer require presentation of the entire stimulus but will restrict stimulus presentation to subregions within the space of a stimulus. These subregions may represent salient regions within a stimulus. The definition and method of selecting these subregions will be discussed in more detail as we discuss the object detection framework.

Chapter 4

Object detection

4.1 Attention

Attention is the ability to economize computational power and reduce its entropy.

4.2 Attention mechanisms

literature reivew of attention mechanisms

4.3 Bottom-up Attention

what we used from Itti's

4.4 Top down attention

describe attempt

Chapter 5

Achieving invariance

5.1 Model

5.2 Results

5.3 Discussion

5.4 Conclusion

Appendix A

The First Appendix

The `\appendix` command should be used only once. Subsequent appendices can be created using the `Chapter` command.

Appendix B

The Second Appendix

Some text for the second Appendix.

Appendix C

Afterword

That's all folks!

C.1 Acknowledgments

Michael Matthew Cook My family: Sahra, father My friends Malte Alf INI