RESEARCH PROJECT IN MECHANICAL <or MECHATRONICS> ENGINEERING

Type of Report

A Convolutional neural network for detecting visual texture

Callan Loomes

Project Report ME110-2022

Co-worker: Conrad Scherb

Supervisor: Dr Luke Hallum

Department of Mechanical and Mechatronics Engineering
The University of Auckland

15 April 2022

ME110-2022

A CONVOLUTIONAL NEURAL NETWORK FOR DETECTING VISUAL TEXTURE

Callan Loomes

DECLARATION

Student

I …….. hereby declare that:

1. This report is the result of the final year project work carried out by my project partner (see cover page) and I under the guidance of our supervisor (see cover page) in the 2021 academic year at the Department of Mechanical Engineering, Faculty of Engineering, University of Auckland.
2. This report is not the outcome of work done previously.
3. This report is not the outcome of work done in collaboration, except that with a project sponsor as stated in the text.
4. This report is not the same as any report, thesis, conference article or journal paper, or any other publication or unpublished work in any format.

In the case of a continuing project: State clearly what has been developed during the project and what was available from previous year(s):

Signature: _____

Date: _____

Supervisor

I confirm that the project work undertaken by this student in the 2022 academic year **is / is not** (*strikethrough as appropriate*) part of a continuing project, components of which have been completed previously.

Comments, if any:

Signature: _____

Date: _____

Table of Contents

Glossary of Terms

Visual Texture	A term referring to the presence of various boundaries/contrasts in images which form the texture of the object being looked at.

Abbreviations

CNN	Convolutional Neural Network
fMRI	Functional magnetic resonance imaging
RF	Receptor Fields
FRF	Filter Rectify Filter
MEG	

1. Introduction

1.1 Project Scope

Visual texture discrimination in humans is a process that is so integrated into modern life, that one scarcely pauses to consider the reason they are able to distinguish the boundary between their wooden cutting board from their marble kitchen bench. This project aims to provide further analysis and understanding behind the mechanisms occurring in the visual cortex when texture boundaries are identified. This will be done through the creation of a CNN that is able to identify different components of texture in an image, distinguishing whether or not there is a boundary between two textures present. The model will be trained on behavioural data generated through human testing, which includes distinguishing whether there is a pattern of orientation in a series of Gabor patches. The scope of the project also encompasses trail and error with different CNN architectures in order to generate the least complex model possible. This is done with the intention that with least complex model will most accurately represent the behaviour of cells in the primary visual cortex. The understanding of texture detection can be applied to a wide range of fields, such as industrial and medical robotics, as well as non-biological autonomous systems, such as self driving vehicles.

1.2 Research Objectives

- Create a CNN which is able to accurately identify whether in a given array of Gabor patches there is a signal present based on orientation.

- Reduce the complexity of our generated CNN model to represent the simplest solution.

- Compare the generated CNN to other computational models and determine its viability as a representative model of how various computational layers in the visual cortex product human image segregation.

- Generate a comprehensive range of testing data which can be used to train various models.

- Investigate the difference in human and CNN responses to recognising signal in a noise array of orientated Gabor patches.

2. Literature Review

The following is a review of the current literature surrounding the use of computational methods to describe the identification of visual textures in humans. It describes the underlying biology behind basic visual identification, as well as the current computational models for describing this identification. An emphasis is then placed on modern CNNs and their use in texture recognition, as these networks are the main focus of the project.

2.1 Underlying Biology Behind Visual Texture Identification

Visual recognition and identification in the human brain is a complex procedure reliant on a hierarchical processing structure of neuron layers in the visual cortex. The use of fMRI has shown that a large amount of neurological processing, with regards to decoding colour and form, is done through the early areas of the visual cortex, V1 to V4 [1]. This early processing includes making simple distinctions, such as orientation or visual texture of perceived objects, which are then conjugated and processed at higher visual layers.

It has been long known that V1, the primary visual cortex is involved with contrast recognition in the brain. In a 1998 study, Tootell et al [2] are able to distinguish the primary visual cortex from other cortical areas through its response to luminance contrast. The same study utilised fMRI and grating stimuli (alternating light and dark patches) to showcase the sensitivity to orientation in human V1. Recent studies have provided a more in depth explanation for this phenomenon. The primary visual cortex consists of two classes of neurones; simple cells and complex cells [3]. Both types of cell contain receptor fields (RF's), regions where light stimuli influence the electrical activity of cells, resulting in light/dark (ON/OFF) regions. Simple cells contain RF's which are spatially specific, and respond strongly to contrasting regions and oriented edges. Complex cells also respond to orientation as well as motion, however are spatially invariant and not as sensitive. A point of contention has been formed with regards to how the simple and complex RF's of the V1 interact to form a retinotopic map which feeds into higher visual layers. Previously it was suggested that RF construction occurred in two consecutive phases, with simple neurons feeding into complex ones [4]. More recent literature, however, suggests a parallel feed-foward system, where the two neuron types are tuned to different stimuli, and combine their RF's concurrently [5]. Due to their sensitivity to orientation and colour contrast, simple cells are fundamental to detecting texture in visual images.

Whilst it is known that basic visual texture identification occurs in the primary visual cortex, the mechanism bridging RF stimulation and pattern identification is sparsely understood. This paves the way for computational models to be used for understanding

how image decoding works in the brain. There is also room for neural network models to explain the conjugative transition from lower level to higher level visual texture recognition [6].

2.2 Previous Models for Cognitive Visual Recognition

A number of computational and machine learning processes have been developed for the purpose of detecting visual texture. These range from a series of rectify models to more machine learning based approaches, each of which have their own strengths, weaknesses and relevance from a biological standpoint.

2.2.1 Non-Machine Learning Approaches

Computational models for attempting to detect visual texture have been around for a while, and a large variety of methods have been tested. In 1991, Bergen and Landy explore a model consisting of subsequent filtering and pooling layers [7]. Many of these models however lack a fundamental biological basis for comparison, meaning they are most likely unrepresentative of the actual process which occurs in the visual cortex [8]. Two types of model however that do have some biological merit are Fourier transfer and FRF approaches.

As early as 1978, V1 neurones were described by Mayhew and Frisby as cells which are responsive to different orientation and spatial frequency combinations [9]. In this study, it was initially theorised that the visual cortex performed a sort of Fourier based transform to discriminate textures, however, after subjects found it easier to discriminate between textures with different spatial frequencies, rather than orientation, this was dismissed. Following studies, however, still suggest the visual cortex performs a Fourier approximation, with particular sensitivities for spatial frequencies [10]. Given the age of these studies and what is now know about the primary visual cortex, it is likely that a purely Fourier based model is unlikely. It is highly possible, however, that some sort of Fourier approximation may still occur at a lower level of analysis in the visual cortex.

Another possible method for segregating visual textures, called a filter rectify filter (FRF) model, is developed by Landy in 2013 [11]. Broken down to its basics, this model first applies a linear filter which is tuned to prefer a particular orientation and spatial frequency; the resulting output when this filter is applied is strongly positive when textures fully match the preferred parameters, strongly negative where textures fully match the opposite to the preferred parameters, and a range in-between. This essentially creates a map where first order, luminance-defined edges are identified. As recognised by Hallum et al and Lian et al, this initial filtering stage draws very similar parallels to how simple and complex cells in the primary visual cortex have receptor fields which respond to orientation and spacial frequency [3,12]. In the subsequent rectify stage, a non linear

threshold or point-wise function separates the positive values (matching alignment) from the negative values (non-matching alignment). Following this, the second and generally larger linear filter is applied which identifies the overarching texture edge in the image. The result of this second filtering is that second order features in the original image are identified. Research suggests the second filtering stage has physiological ties to the V2, V3 and V4 areas of the visual cortex, with evidence that it may begin as early as the V1 layer [12, 13].

FRF and Fourier computational methods for segmenting visual textures are useful for generally describing the low level processes that occur in the visual cortex. They do, however, leave room for more complicated models to explain these processes in more detail, and link their use in higher level visual identification.

2.2.2 Machine Learning Approaches

In recent literature, the use of machine learning models such as deep neural networks and CNNs has been a very popular and effective technique for mapping and understanding regions of the brain. In 2016, Cichy et al created a deep neural network consisting of eight layers, each of which performed a variety of tasks such as convolution, pooling or normalisation [14]. While previous studies relied on extrapolating data from well understood areas such as the primary visual cortex to generate behaviour, this neural network built these pathways from scratch against large sets of training data. When compared with real data generated in space and time via fMRI and MEG respectively, both the trained model and real data showed a hierarchy where lower level processing was conjugated into higher level processing. Hence, studying the CNN architecture potentially provides insight into similar architecture in the brain. Interestingly, the hierarchical structure of the CNN only manifested after it was trained with data, and was not based on the architecture of the network.

Machine learning has also been used in similar applications such as the decoding and visualisation of EEG signals from the brain [15]. In this study, a range of network architectures were tested on raw data to generate a final CNN which was able to detect patterns in EEG signals. More regarding the individual architectures of these networks is discussed in section 2.3, however, the different testing performed in this paper highlighted the importance of both network architecture and training data on the performance of these networks.

2.2.3 Previous Relevant testing Data

The training data presented to a machine learning model is one of the key factors that determines its success. Training data need to be sufficiently consistent, accurate and in a form where it can be easily understood by machine learning algorithms. Additionally, there needs to be a sufficient amount of training data available, excess to the capacity

(number of training parameters) of the neural network to ensure memorization does not occur [16]. A training parameter which fits this mould are Gabor patches. Gabor patches can be easily generated in a variety of different orientations and sizes through computational convolution, and can be easily detected and analysed by computational models. Furthermore, Gabor patches relate strongly to this study, as their simplistic nature means that most of the detection and analysis of their position and orientation occurs through the primary visual cortex; studies have shown that Gabor patches can be used to stimulate the V1 to configurations commonly seen in nature [17]. Gabor patches have been used as an experimental medium in many studies, such as in Claessens and Wagemans (2005) where the effect of orientation and proximity of a group of Gabor patches is investigated with regard to visual stimuli [18]. Gabor patches have also been used in clinical studies, particularly in triggering contrast responses in the primary visual cortex for patients with low myopia or early presbyopia [19].

Another topic of relevance to this study refers to the human ability to detect patterns from from noise and vice versa. This is particularly important if behavioural training data used relies on distinguishing images with no orientation pattern (i.e. random noise) with those where a pattern is present. In very early studies, Burgess et al. (1981) suggest that humans have a high efficient for the separation of signal from noise, at around 70% efficiency of detection [20]. More recent studies show a similar trend, however also note that the quality of signal detection increases after an increased amount of training [21]. This may be an important factor when designing trials for the generation of training data.

2.3 Current Literature regarding CNNs

Given the promising nature of CNNs for visual texture recognition, their use to further explore identification of visual texture is warranted. A convolution network is one where at least one or more of its layers involve a convolution step [15]. The benefit of these convolution steps are their ability for highly specific edge detection by breaking down images into texture components. Traditionally, convolutional neural networks were designed on the principles of biology, and hence are very suitable for the task of modelling the visual cortex [22].

2.3.1 Overall Architecture and Performance

The three main points of difference that separate different neural networks are the overall architecture, the training procedure and the task to be learned. Of these three, neural architecture is the most fundamental, and has a large influence on the performance and efficiency of the network. There are many studies which give methods of designing and hyper tuning CNNs to improve performance. In general, deep convolutional networks consist of a collection of layers which feed forward into each other [23]. Each layer of the system has a different function, which can range from convolution,

noise reduction, pooling, correction layers and more [14]. Each layer has specific gains and parameters that can be changed, and different weightings with regards to feeding into subsequent layers; these are the parameters that give the network the ability to 'learn' through back-propagation of data, as there parameters are tuned such that the networks responds in a particular way to input stimuli.

Schirrmeister et al explore a large range of architecture design considerations in their 2017 study regarding the decoding of EEGs [15]. Deep ConvNets were designed which extracted a large number of generic features, rather than specific ones. These deep CNNs were characterised by their large number of pooling blocks and classification layers. To contrast, shallow ConvNets were more focused on specific feature extraction, and its architecture consisted more of processing convolution and filter layers, which were then followed by a few pooling and soft-max layers. Hybrid deep-shallow CNNs were also investigated, which combined both generic and specific feature extraction. From this study, the link between architecture and function is highlighted; by focusing on different types of layers, different output characteristics could be realised.

Another case study of CNN architecture is the network designed for ImageNet classification [24]. In this example, a ConvNet is generated with five convolutional layers, followed by three pooling layers, three fully connected layers and lastly a 1000-way Softmax. Of these layers, the convolutional and fully connected layers are weighted/trained, with the pooling layers generating links between the learned layers, and the soft max layer applying classes to the final computational outputs. Upon removing one of the convolutional layers, accuracy of the ConvNet degraded significantly, and hence this resembled a minimum cost architecture.

2.3.2 Individual Layer Design

While overall architecture highlights how different layers come together to form the decision making of the network, individual layer design is extremely important to ensure the big picture performs. There are a large amount of layer types available, each with a different emphasis on computation, decision making, and pattern recognition. Sultana et al highlight a large number of possible layers in their 2018 study [25]. Convolutional layers provide the core computational breakdown of images; they use a kernel which passes over the input image and performs computational convolution on each position. The kernel size is a parameter which can be adjusted, with larger kernels increasing computational time for a decrease in spatial specificity [26]. Pooling layers essentially break down larger images into smaller ones to reduce computation effort by sampling or averaging regions in an image. Fully connected layers are the basic decision making layers, as they take inputs from each previous layer and produce an output based on the results of those layers. Other classification and filtering layers also exist, such as Softmax and low-pass filter layers respectively, which can further help with processing

images.

2.3.3 Resources available for for building

One of the most computationally expensive steps of CNN construction is the back-propagation training required. Fortunately, recent studies have focused on this issue, and generated automatic computational techniques to ensure models are not over constrained with training data [27]. The theory behind this computational technique is to use a low cost strategy that predicts the of a CNN after a given amount of training data is inputted. This means the number of approximate training iterations can be found which provides optimal performance. The benefit of this is that it not only reduces training time, but also the risk of over constraining the system, meaning it relies more on memory than adaptive learning. A similar method to reduce over fitting is also used in [24].

References

[1] J. Taylor and Y. Xu, "Representation of color, form, and their conjunction across the human ventral visual pathway," *NeuroImage*, vol. 251, p. 118941, may 2022.

[2] R. B. H. Tootell, N. K. Hadjikhani, W. Vanduffel, A. K. Liu, J. D. Mendola, M. I. Sereno, and A. M. Dale, "Functional analysis of primary visual cortex (v1) in humans," *Proceedings of the National Academy of Sciences*, vol. 95, no. 3, pp. 811–817, feb 1998.

[3] Y. Lian, A. Almasi, D. B. Grayden, T. Kameneva, A. N. Burkitt, and H. Meffin, "Learning receptive field properties of complex cells in v1," *PLOS Computational Biology*, vol. 17, no. 3, p. e1007957, mar 2021.

[4] L. M. Martinez and J.-M. Alonso, "Complex receptive fields in primary visual cortex," *The Neuroscientist*, vol. 9, no. 5, pp. 317–331, oct 2003.

[5] G. Kim, J. Jang, and S.-B. Paik, "Periodic clustering of simple and complex cells in visual cortex," *Neural Networks*, vol. 143, pp. 148–160, nov 2021.

[6] E. Peterhans and R. von der Heydt, "Subjective contours - bridging the gap between psychophysics and physiology," *Trends in Neurosciences*, vol. 14, no. 3, pp. 112–119, mar 1991.

[7] J. R. Bergen and M. S. Landy, "Computational modeling of visual texture segregation," in *Computational Models of Visual Processing*. The MIT Press, 1991.

[8] H. R. Wilson, "Non-fourier cortical processes in texture, form, and motion perception," in *Cerebral Cortex*. Springer US, 1999, pp. 445–477.

[9] J. E. W. MAYHEW and J. P. FRISBY, "Texture discrimination and fourier analysis in human vision," *Nature*, vol. 275, no. 5679, pp. 438–439, oct 1978.

[10] F. L. Royer, M. S. Rzeszotarski, and G. C. Gilmore, "Application of two-dimensional fourier transforms to problems of visual perception," *Behavior Research Methods and Instrumentation*, vol. 15, no. 2, pp. 319–326, mar 1983.

[11] M. S. Landy, *The New Visual Neurosciences*, 2013, ch. Texture analysis and perception, pp. 639–652.

[12] L. E. Hallum, M. S. Landy, and D. J. Heeger, "Human primary visual cortex (v1) is selective for second-order spatial frequency," *Journal of Neurophysiology*, vol. 105, no. 5, pp. 2121–2131, may 2011.

[13] B. A. Wandell, S. O. Dumoulin, and A. A. Brewer, "Visual field maps in human cortex," *Neuron*, vol. 56, no. 2, pp. 366–383, oct 2007.

[14] R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, and A. Oliva, "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence," *Scientific Reports*, vol. 6, no. 1, jun 2016.

[15] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, aug 2017.

[16] D. Arpit, S. Jastrzebski, N. Ballas, D. Krueger, E. Bengio, M. S. Kanwal, T. Maharaj, A. Fischer, A. Courville, Y. Bengio, and S. Lacoste-Julien, "A closer look at memorization in deep networks," 2017.

[17] U. Ernst, S. Denève, and G. Meinhardt, "Detection of gabor patch arrangements is explained by natural image statistics," *BMC Neuroscience*, vol. 8, no. S2, jul 2007.

[18] P. M. Claessens and J. Wagemans, "Perceptual grouping in gabor lattices: Proximity and alignment," *Perception and Psychophysics*, vol. 67, no. 8, pp. 1446–1459, nov 2005.

[19] D. Durrie and P. S. McMinn, "Computer-based primary visual cortex training for treatment of low myopia and early presbyopia." *Transactions of the American Ophthalmological Society*, vol. 105, pp. 132–8; discussion 138–40, 2007.

[20] A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," *Science*, vol. 214, no. 4516, pp. 93–94, oct 1981.

[21] Z. Hussain and P. J. Bennett, "Perceptual learning of detection of textures in noise," *Journal of Vision*, vol. 20, no. 7, p. 22, jul 2020.

[22] J. M. Vaz and S. Balaji, "Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics," *Molecular Diversity*, vol. 25, no. 3, pp. 1569–1584, may 2021.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, may 2015.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, may 2017.

[25] F. Sultana, A. Sufian, and P. Dutta, "Advancements in image classification using convolutional neural network," in *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, nov 2018.

[26] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, jun 2018.

[27] G. Franchini, , V. Ruggiero, F. Porta, and L. Z. and, "Neural architecture search via standard machine learning methodologies," *Mathematics in Engineering*, vol. 5, no. 1, pp. 1–21, 2022.