A Convolutional Neural Network for Detecting Visual Texture

Callan Loomes, Conrad Scherb

Supervisor: Dr. Luke Hallum

UNIVERSITY OF AUCKLAND Waipapa Taumata Rau NEW ZEALAND

ENGINEERING
DEPARTMENT OF MECHANICAL AND MECHATRONICS ENGINEERING

Background and Scope

Humans recognize boundaries through both **luminescence and texture** differences; the latter of the two is dependent on orientation of a surface, and is the focus of this study.

Texture is processed in the **Primary Visual Cortex** (V1); neurons highlight vision regions called **receptive fields** (RFs), which are sensitive to the spatial frequency and orientation of a surface (*Figure 1*) [1]. Current literature suggests that we can use **convolutional neural networks** (CNNs) to model texture processing in the V1 [2].

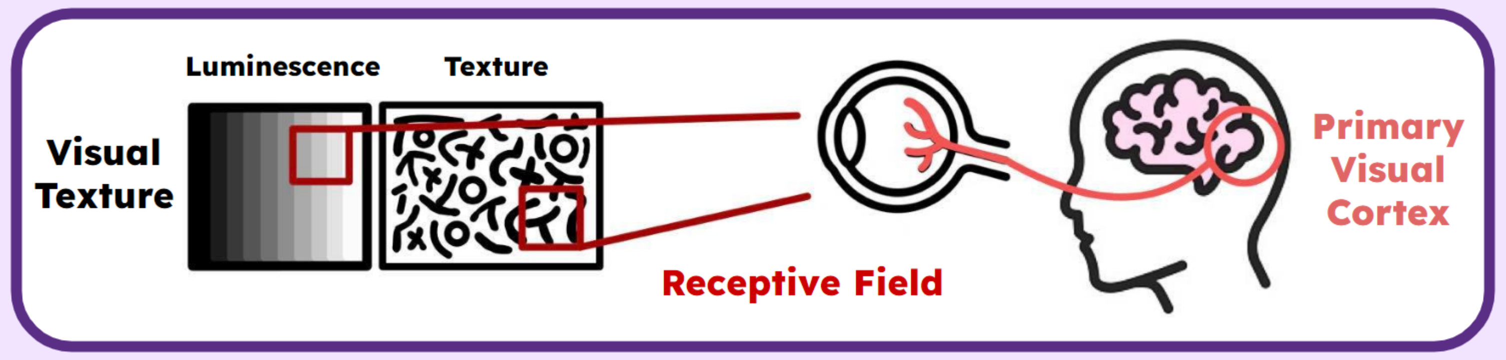

Figure 1: *Pathway of texture visualization in humans*

Aims:
- Create data for human response to visual texture
- Vary CNN structure to match human responses
- Link CNN models to existing biological models

1 Dataset Generation and Data Collection

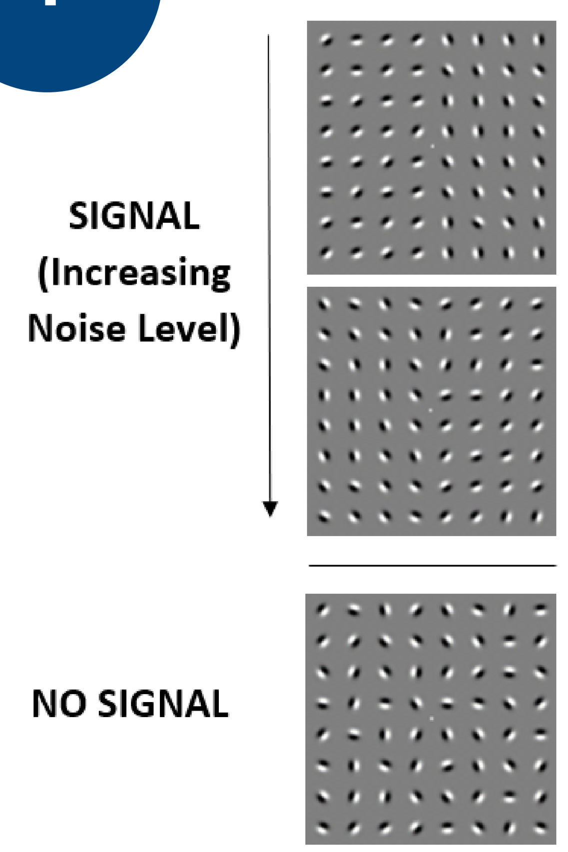

Figure 2: *Datasets with signal vs no signal*

Visual texture was controlled using an array of **Gabor Patches.** These are common in visual studies due to their pronounced orientation [2].

Images with a texture signal have a vertical boundary down the middle of the image, while those with no signal have random orientations (*Figure 2*). Datasets of 2500 images (50% signal) were created for testing small, medium and large textures (*Figure 3*). Participants had accuracies of 85.5%, 85.3% and 88.2% respectively for the datasets.

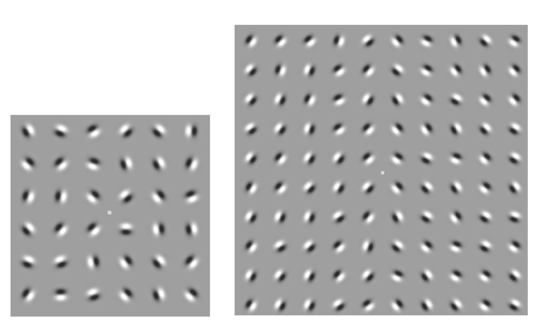

Figure 3: *Smaller vs larger dataset*

Data collection Process

1) Image flashed for 0.5s, such that judgement is based on only the V1.

2) User inputs if a signal is present

3) Process repeated for the next random dataset

2 Model Training and Parameters

- Tensorflow, with the Keras API and **ten fold cross-validation**, were used to train models.
- The following hyperparameters were varied from a base model (*Figure 4*):
  - Kernel Size
  - Number of Neurons
  - Number of convolutional layers
- Trained for 350 epochs and until loss plateaued.

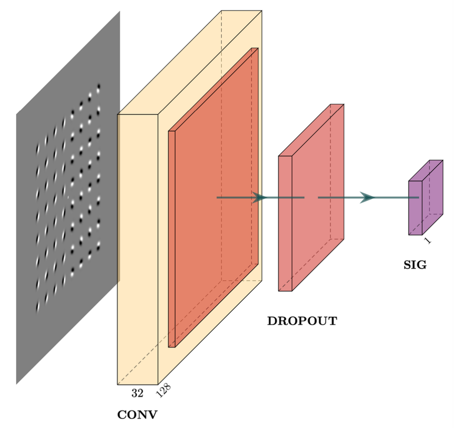

Figure 4: *Base CNN structure*

Results

Changing Kernel Size

- Kernel size represents the area of the convolution filters.
- Filters that are too large or small cannot detect patterns between the Gabor patches.

Neuron No: 32 (all models)

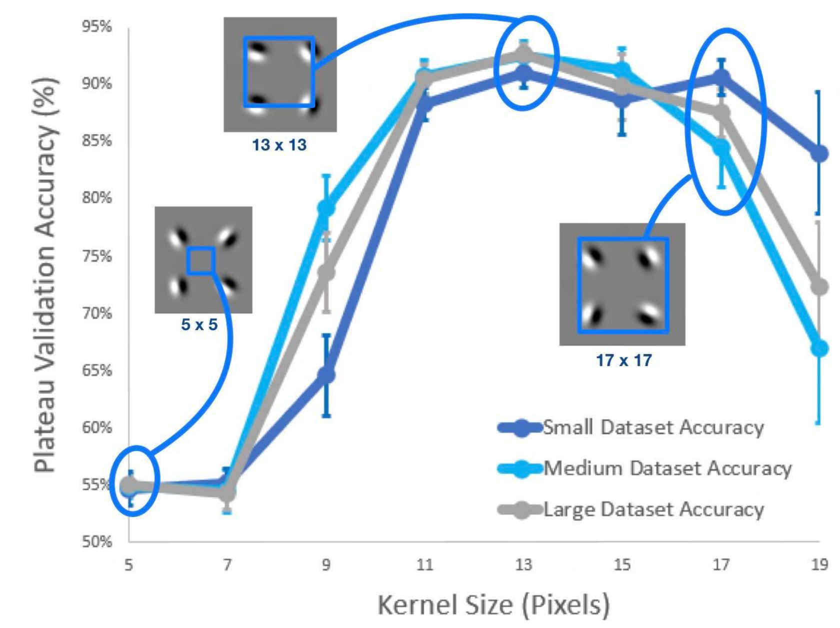

Figure 5: *Graph showing relationship between kernel size and accuracy for single layer CNN with 32 neurons.*

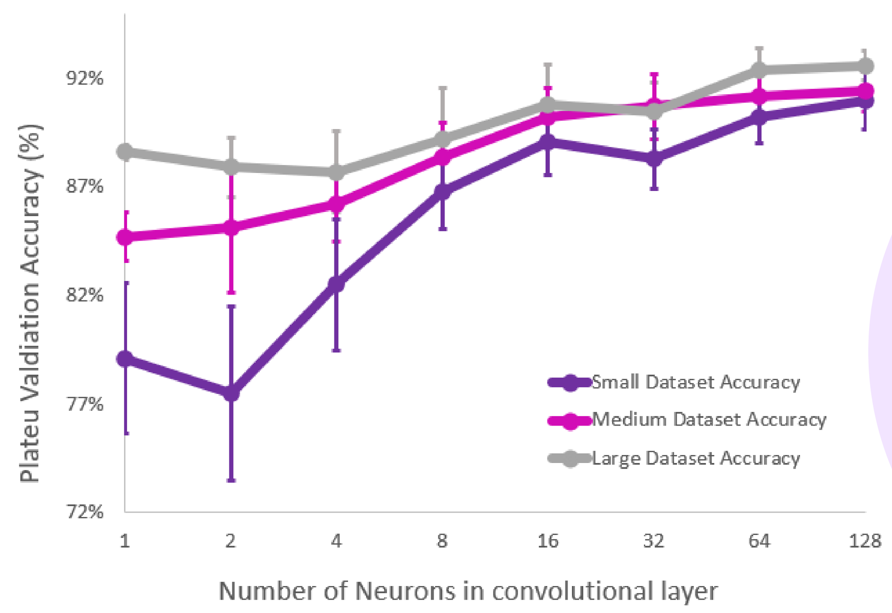

Figure 6: *Graph showing relationship between number of neurons and accuracy for single layer CNN with 11x11 kernel size*

Changing Neuron Number

- The number of neurons represents the number of different convolutional filters.
- Greater filter number means more distinguishing features can be identified.

Kernel Size: 11 (all models)

Increased Model Depth

- Adding more layers could allow the network to learn more integrated information
- Too many convolutional layers can cause overfitting and data that is too convoluted.

Neuron No: (1st) 32, 16, 8, 8, 8
Kernel Size: (1st) 13, 3, 3, 3, 3

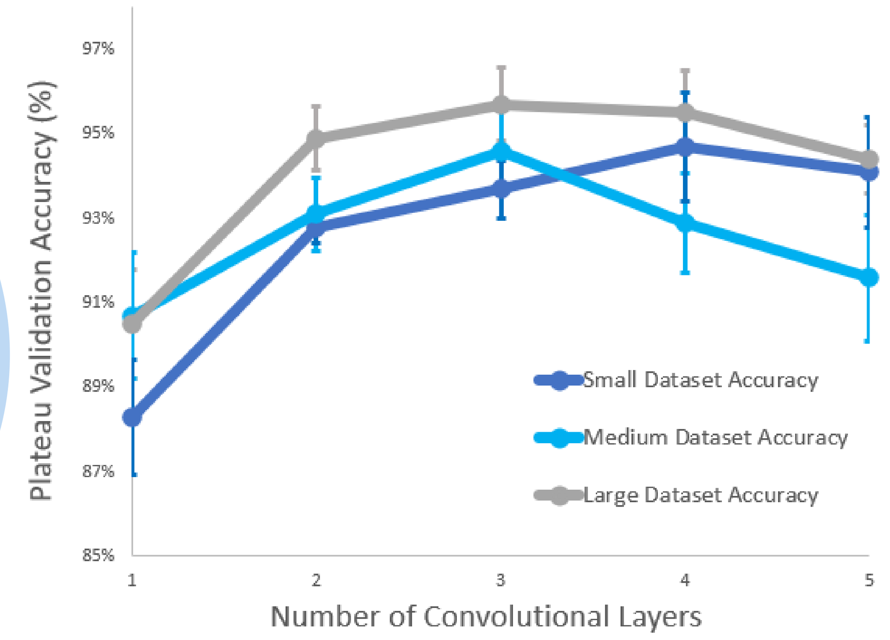

Figure 7: *Graph showing relationship between number of conv layers and accuracy.*

3 Biological Implications of Models

Results show that the highest accuracy occurs at higher neuron numbers, and a kernel size of 13x13 pixels. This trend is consistent for different combinations of hyper-parameters (*Figure 8*). The following biological inferences can be deduced from this data:

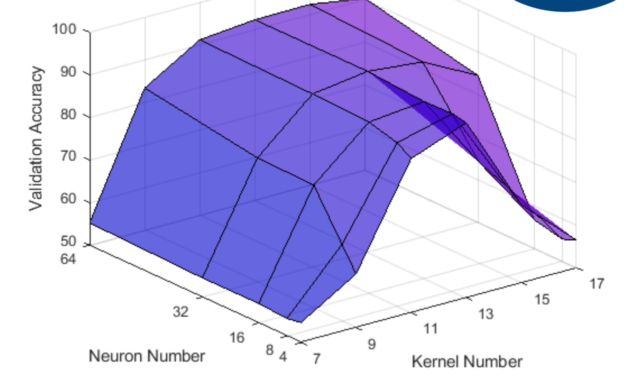

Figure 8: *Accuracy of models for varied kernel size and neuron numbers.*

1) Kernels directly model receptive fields [3], hence V1 neuron RFs may have the same size ratio as the 13x13 pixels in this study.

2) Increasing neuron number increases accuracy. This may support existing literature which suggests each neuron is specific to a particular orientation [4].

3) The most efficient number of interconnected neuron layers is around 2-4. This result is similar to existing literature [5].

Further to **(1)** and **(2)**, the output kernels from trained models show **directional selectivity** (*Figure 9*). Combined with the single layered CNNs, this indicates a V1 structure that is similar to the **filter-rectify-filter** (FRF) model suggested by Hallum et. al [4].

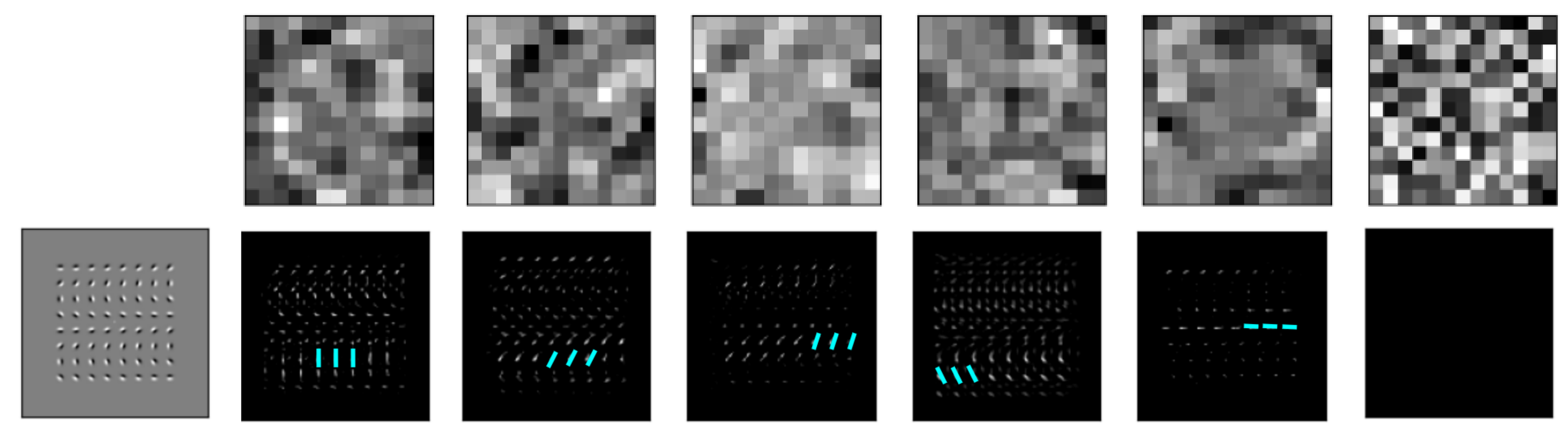

Figure 9: *Kernel output from an 8-neuron, 13x13 kernel size, single layered model, convoluted against a test image containing Gabor patches orientated at all angles (patterns highlighted).*

Conclusions and Future Research

- **Receptor fields** of V1 neurons may span multiple texture units.
- V1 texture detection may occur via 2-4 interconnected layers.
- Single layered models and **orientation selective kernels** provide evidence to support the **FRF model** of the V1.

Future Research:

- Scale up the project to include more complex textures.
- Analyse EEG during the testing process, and perform testing on a larger and more diverse population.

References

[1] M. S. Landy, The New Visual Neurosciences, 2013, ch. Texture analysis and perception, pp. 639–652. **[2]** Laskar, M. N. U., Sanchez Giraldo, L. G., & Schwartz, O. Deep neural networks capture texture sensitivity in V2 ,2020. *Journal of Vision*, *20*(7), 21. **[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. 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. **[5]** Kociołek, M., Kozłowski, M., & Cardone, A. A Convolutional Neural Networks-Based Approach for Texture Directionality Detection, 2022. Sensors, 22(2), 562.