Maximise Activations

VLC = ∭ Vision Learning Control dVdLdC

Visualisation

Ethan Harris

Vision, Learning and Control
University of Southampton

Overview

- The Electrophysical and Psychophysical Aspects of Vision
- Characterising Single Cells
- Feature Visualisation: Monkeys to Machines
  - Decorrelation
  - Reparameterisation
  - Maximising: Cells, Layers, Predictions
- Machines to Monkeys

Note

- Feature visualisation article here: `https://distill.pub/2017/feature-visualization/`
- PyTorch implementations of key algorithms which generated the images can be found here: `https://github.com/pytorchbearer/visual`

The Electrophysical and Psychophysical Aspects of Vision

- We want to know the function of the brains input that gives rise to some output

The Electrophysical and Psychophysical Aspects of Vision

- We want to know the function of the brains input that gives rise to some output
- Inputs - visual stimuli

The Electrophysical and Psychophysical Aspects of Vision

- We want to know the function of the brains input that gives rise to some output
- Inputs - visual stimuli
  - still images, moving stimuli, colour, greyscale, ...

The Electrophysical and Psychophysical Aspects of Vision

- We want to know the function of the brains input that gives rise to some output
- Inputs - visual stimuli
  - still images, moving stimuli, colour, greyscale, ...
- Outputs - activations

The Electrophysical and Psychophysical Aspects of Vision

- We want to know the function of the brains input that gives rise to some output
- Inputs - visual stimuli
  - still images, moving stimuli, colour, greyscale, ...
- Outputs - activations
  - single cells, groups of cells, external actions, micro-electrode recordings, fMRI, ...

Characterising Single Cells

- Using a micro-electrode recording from a cell, we can plot the **response curve** to a range of stimuli

Characterising Single Cells

- Using a micro-electrode recording from a cell, we can plot the **response curve** to a range of stimuli
- Need to restrict to some controlled stimuli space - colour? edge orientation?

Case Study: Colour Opponency

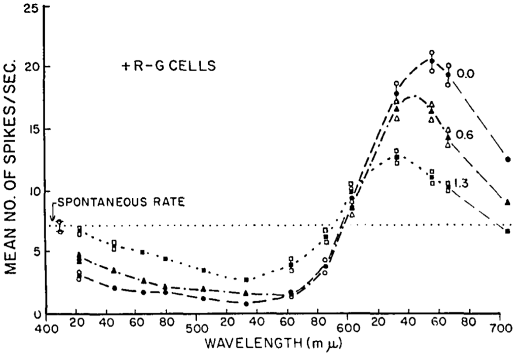
- Consider the baseline response (or spontaneous rate) of a cell to an empty visual field (i.e. constant grey-level across the retina)

Case Study: Colour Opponency

- Consider the baseline response (or spontaneous rate) of a cell to an empty visual field (i.e. constant grey-level across the retina)
- Now, we change the colour (wavelength) of the visual stimuli and measure the response

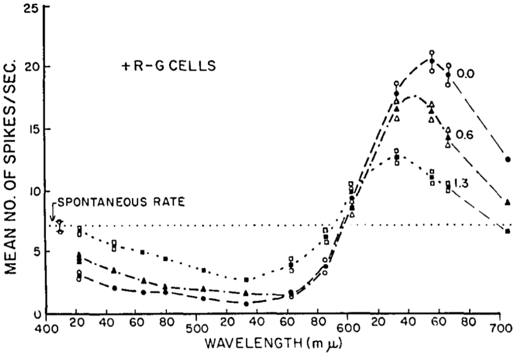
De Valois: Analysis of Response Patterns of LGN Cells

- Macaque Lateral Geniculate Nucleus (LGN) cells



De Valois: Analysis of Response Patterns of LGN Cells

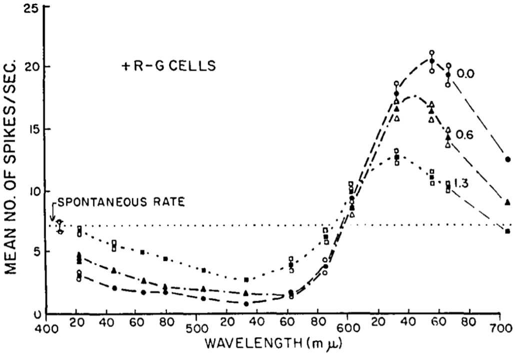
- Macaque Lateral Geniculate Nucleus (LGN) cells



- Opponent if the curve crosses the line - Non-opponent otherwise

De Valois: Analysis of Response Patterns of LGN Cells

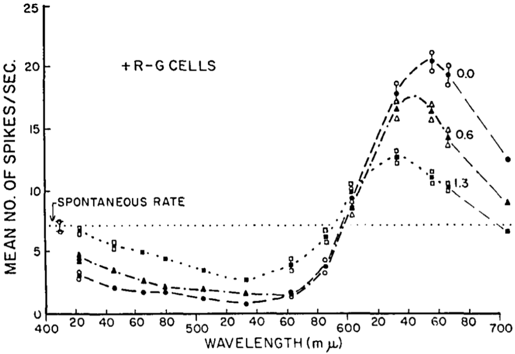
- Macaque Lateral Geniculate Nucleus (LGN) cells



- Opponent if the curve crosses the line - Non-opponent otherwise
- An analogous form of opponency can be defined - **spatial** opponency

De Valois: Analysis of Response Patterns of LGN Cells

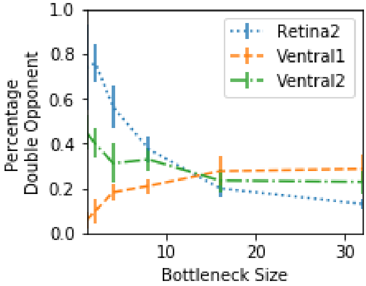
- Macaque Lateral Geniculate Nucleus (LGN) cells



- Opponent if the curve crosses the line - Non-opponent otherwise
- An analogous form of opponency can be defined - **spatial** opponency
- Cells which are both spatially opponent and colour opponent are called **double** opponent

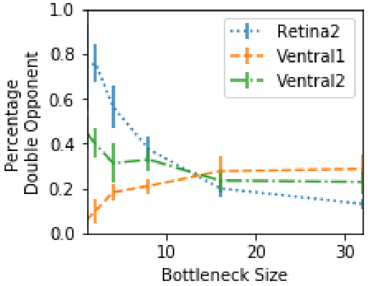
Opponency in Deep Networks

- Turns out both spatial and colour opponency happen in deep networks too!



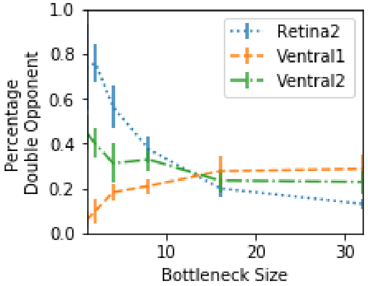
Opponency in Deep Networks

- Turns out both spatial and colour opponency happen in deep networks too!
- We measure in **exactly** the same way as we would in a Monkey - but on a much bigger scale



Opponency in Deep Networks

- Turns out both spatial and colour opponency happen in deep networks too!
- We measure in **exactly** the same way as we would in a Monkey - but on a much bigger scale
- More here: `https://github.com/ecs-vlc/opponency`



Feature Visualisation: Monkeys to Machines

- We can also ask the question 'what most excites this cell?' in a deep network

Feature Visualisation: Monkeys to Machines

- We can also ask the question 'what most excites this cell?' in a deep network
- We've shown a range of stimuli and plotted a response curve as with the monkey

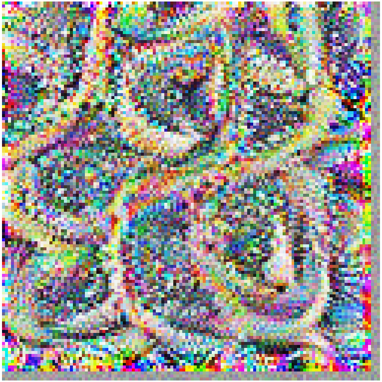
Feature Visualisation: Monkeys to Machines

- We can also ask the question 'what most excites this cell?' in a deep network
- We've shown a range of stimuli and plotted a response curve as with the monkey
- Can we do something smarter?

Feature Visualisation: Monkeys to Machines

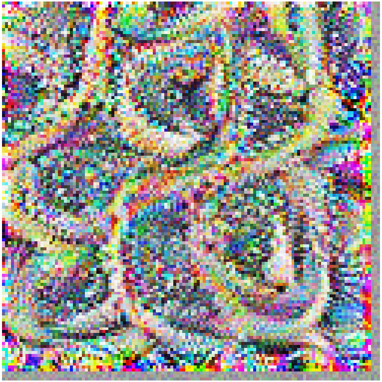
- We can also ask the question 'what most excites this cell?' in a deep network
- We've shown a range of stimuli and plotted a response curve as with the monkey
- Can we do something smarter?
- Gradient ascent - use the gradient information we have to 'learn' the image which maximises the response of a particular cell or channel

Gradient Ascent



- Inception V3, Layer 6 - Noisy image - not very informative

Gradient Ascent



- Inception V3, Layer 6 - Noisy image - not very informative
- Gradient ascent makes an **uncorrelated** update in a **correlated** space - hence the crazy colours

Colour Correlation

- Compute the RGB correlation statistics from ImageNet

Colour Correlation

- Compute the RGB correlation statistics from ImageNet
- Correlate the colour channels of the input according to these statistics at each step - now the gradient update is over the **uncorrelated** parameters

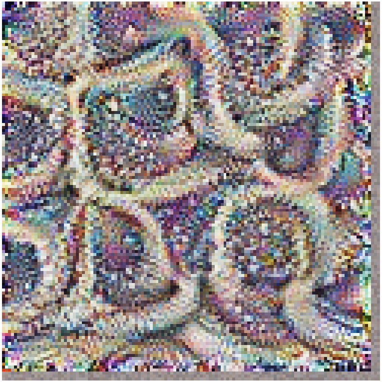
Colour Correlation

- Compute the RGB correlation statistics from ImageNet
- Correlate the colour channels of the input according to these statistics at each step - now the gradient update is over the **uncorrelated** parameters
- This type of trick is referred to as **re-paramterisation** or **preconditioning**

Colour Correlation

- Compute the RGB correlation statistics from ImageNet
- Correlate the colour channels of the input according to these statistics at each step - now the gradient update is over the **uncorrelated** parameters
- This type of trick is referred to as **re-paramterisation** or **preconditioning**
- The maxima don't change - but the loss surface does - some optima become more likely to be found

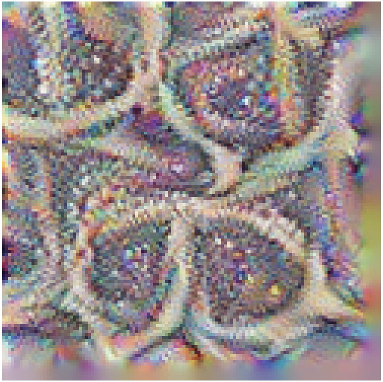
Colour Correlation



- What can we do about the noise?

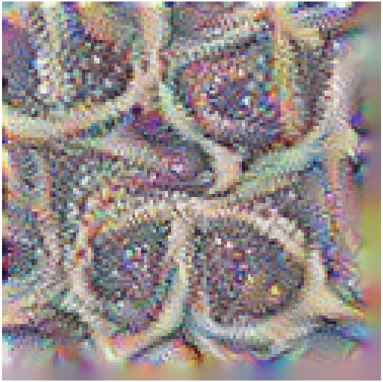
Frequency Penalisation

- We can construct additional loss functions that penalise high frequencies
  - Total variation, blur, L1 loss, ...



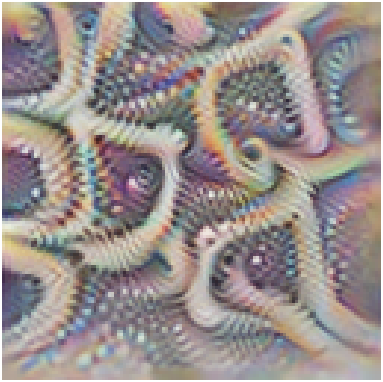
Frequency Penalisation

- We can construct additional loss functions that penalise high frequencies
  - Total variation, blur, L1 loss, ...
- It's better - but we have other ways to improve optimisation procedures



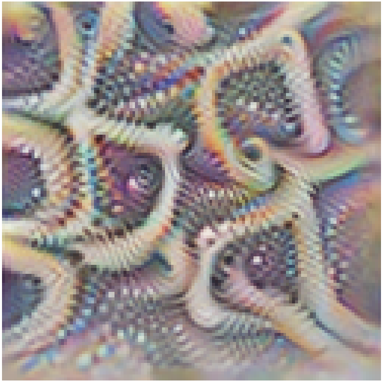
Augmentation

- Randomly transform our image before each step
- parameters $\rightarrow$ correlation $\rightarrow$ frequency penalisation $\rightarrow$ augmentation $\rightarrow$ model



Augmentation

- Randomly transform our image before each step
- parameters $\rightarrow$ correlation $\rightarrow$ frequency penalisation $\rightarrow$ augmentation $\rightarrow$ model
- We know the importance of correlation - what else is correlated in natural images?



Fourier Space

- Space! pixel values are correlated with their neighbours

Fourier Space

- Space! pixel values are correlated with their neighbours
- But how do we model spatial correlation?

Fourier Space

- Space! pixel values are correlated with their neighbours
- But how do we model spatial correlation?
- If a correlation is spatially consistent (i.e. it is constant over the extent of the image) then the Fourier coefficients are **independent**

Fourier Space

- Space! pixel values are correlated with their neighbours
- But how do we model spatial correlation?
- If a correlation is spatially consistent (i.e. it is constant over the extent of the image) then the Fourier coefficients are **independent**
- Think of a spatially consistent correlation as a convolution operation

Fourier Space

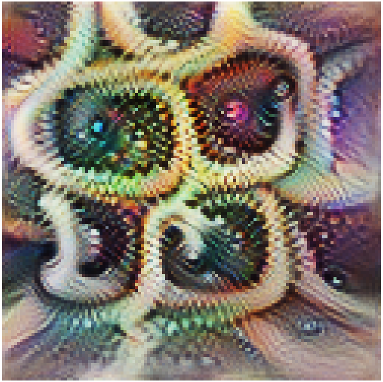
- Space! pixel values are correlated with their neighbours
- But how do we model spatial correlation?
- If a correlation is spatially consistent (i.e. it is constant over the extent of the image) then the Fourier coefficients are **independent**
- Think of a spatially consistent correlation as a convolution operation
- By the convolution theorem, this is a point-wise multiplication in Fourier space - it treats the coefficients independently

Fourier Space

- Space! pixel values are correlated with their neighbours
- But how do we model spatial correlation?
- If a correlation is spatially consistent (i.e. it is constant over the extent of the image) then the Fourier coefficients are **independent**
- Think of a spatially consistent correlation as a convolution operation
- By the convolution theorem, this is a point-wise multiplication in Fourier space - it treats the coefficients independently
- The Fast Fourier Transform (FFT) is differentiable!

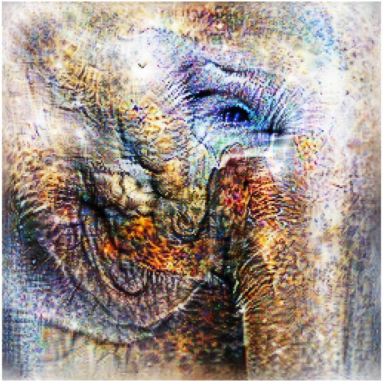
Fourier Space

- Fourier coefficients $\rightarrow$ parameters $\rightarrow$ correlation $\rightarrow$ augmentation $\rightarrow$ model - frequency penalisation is turned off here
- That's better!



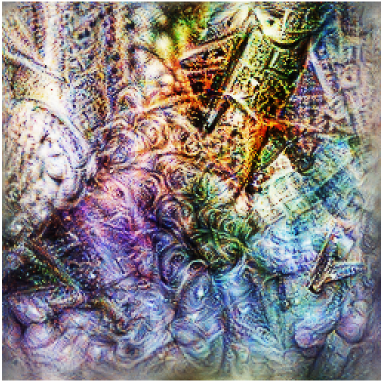
Maximising the Outputs

- We could also maximise the outputs for a particular class
- Here's an Indian Elephant (class 385)



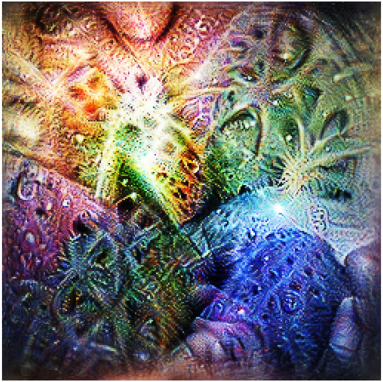
Maximising the Outputs

- And a Grasshopper (class 311)



Maximising the Outputs

- Or a Strawberry (class 949)
- You get the idea - deep neural networks don't learn about shape!



Maximising the Interestingness

- What if we maximise the sum of the squares of the outputs from a whole layer?

Maximising the Interestingness

- What if we maximise the sum of the squares of the outputs from a whole layer?
- DeepDream! - best results when we start with a real image and gradually increase the scale through training

Mona Lisa

Mona Lisa

And Back to Monkeys

- So we've seen how we can get great visualisations of deep networks

And Back to Monkeys

- So we've seen how we can get great visualisations of deep networks
- Much more informative than just looking at single cells

And Back to Monkeys

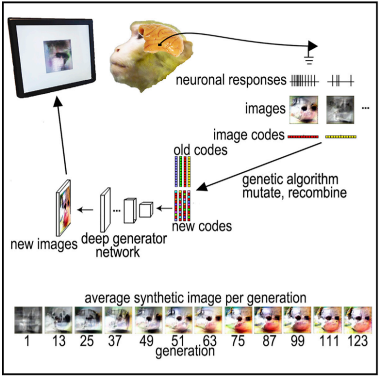
- So we've seen how we can get great visualisations of deep networks
- Much more informative than just looking at single cells
- What's stopping us from doing this in a real brain?

And Back to Monkeys

- So we've seen how we can get great visualisations of deep networks
- Much more informative than just looking at single cells
- What's stopping us from doing this in a real brain?
- Nothing! - gradient free optimisers exist - we could use a genetic algorithm

XDREAM

- Monkey + Genetic Algorithm + Deep Generative Network



XDREAM

- Neurons in the Inferior Temporal cortex (AKA the 'what' pathway)

Closing Remarks

- Finding the stimuli which maximise the response of single cells has a long history in Neuroscience

Closing Remarks

- Finding the stimuli which maximise the response of single cells has a long history in Neuroscience
- Applying that idea to deep learning we get a way to understand what each 'cell' has learned

Closing Remarks

- Finding the stimuli which maximise the response of single cells has a long history in Neuroscience
- Applying that idea to deep learning we get a way to understand what each 'cell' has learned
- But be careful - feature visualisation is more of an art than a science - it isn't a completely solved problem

Closing Remarks

- Finding the stimuli which maximise the response of single cells has a long history in Neuroscience
- Applying that idea to deep learning we get a way to understand what each ‘cell’ has learned
- But be careful - feature visualisation is more of an art than a science - it isn’t a completely solved problem
- Lessons and tools from deep learning can feed back in to Neuroscience, helping us to understand **much** more complex parts of the brain

Closing Remarks

- Finding the stimuli which maximise the response of single cells has a long history in Neuroscience
- Applying that idea to deep learning we get a way to understand what each ‘cell’ has learned
- But be careful - feature visualisation is more of an art than a science - it isn’t a completely solved problem
- Lessons and tools from deep learning can feed back in to Neuroscience, helping us to understand **much** more complex parts of the brain
- Maybe deep models and brains aren’t so different after all!