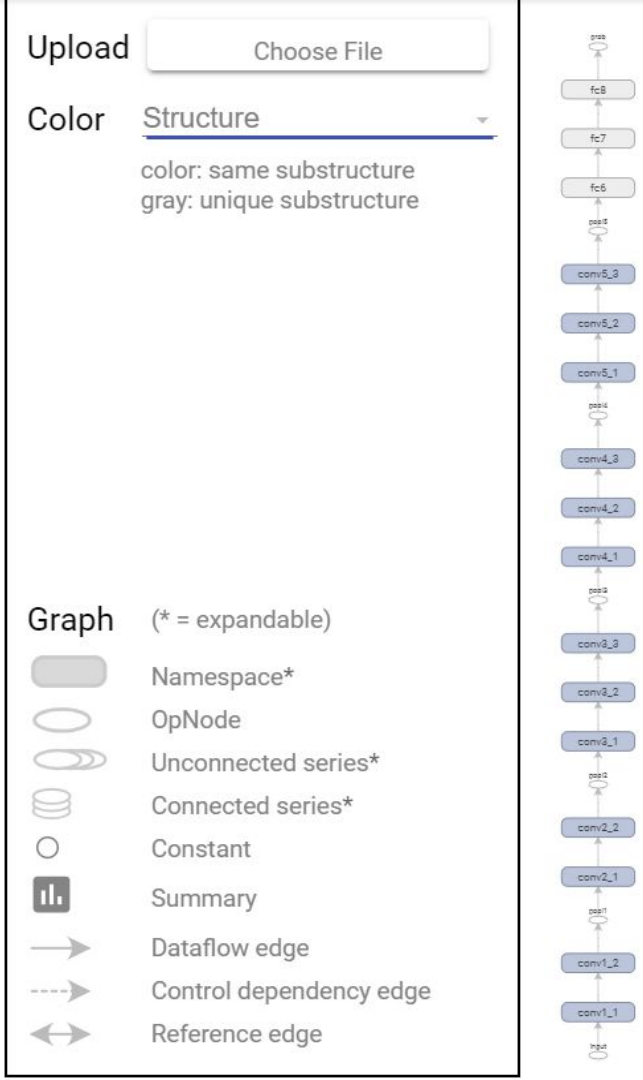
VGG16_caffe model

Application of Google lucid visualisations

VGG Model framework

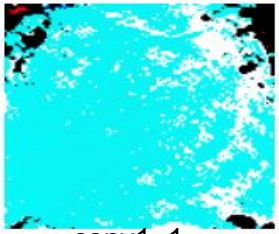
Model chosen here was VGG16 model as this had comparatively less layers.

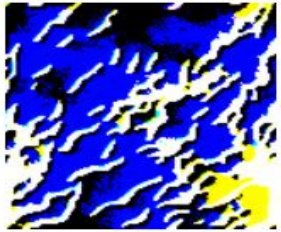
{'name': 'conv1_1/conv1_1', 'size': 64, 'type': 'conv'},
{'name': 'conv1_2/conv1_2', 'size': 64, 'type': 'conv'},
{'name': 'conv2_1/conv2_1', 'size': 128, 'type': 'conv'},
{'name': 'conv2_2/conv2_2', 'size': 128, 'type': 'conv'},
{'name': 'conv3_1/conv3_1', 'size': 256, 'type': 'conv'},
{'name': 'conv3_2/conv3_2', 'size': 256, 'type': 'conv'},
{'name': 'conv3_3/conv3_3', 'size': 256, 'type': 'conv'},
{'name': 'conv4_1/conv4_1', 'size': 512, 'type': 'conv'},
{'name': 'conv4_2/conv4_2', 'size': 512, 'type': 'conv'},
{'name': 'conv4_3/conv4_3', 'size': 512, 'type': 'conv'},
{'name': 'conv5_1/conv5_1', 'size': 512, 'type': 'conv'},
{'name': 'conv5_2/conv5_2', 'size': 512, 'type': 'conv'},
{'name': 'conv5_3/conv5_3', 'size': 512, 'type': 'conv'},
{'name': 'fc6/fc6', 'size': 4096, 'type': 'dense'},
{'name': 'fc7/fc7', 'size': 4096, 'type': 'dense'},
{'name': 'prob', 'size': 1000, 'type': 'dense'}

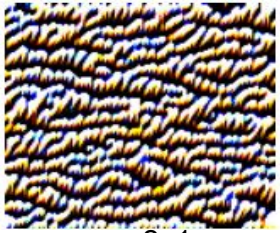


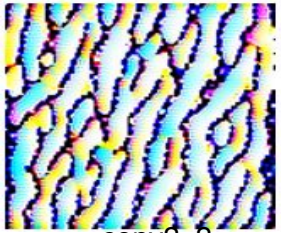
Neuron visualisation

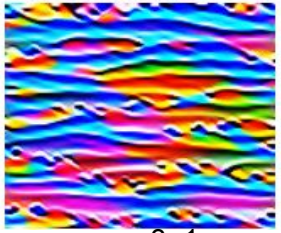
Examples of neurons from each layer *(unit 1 from each layer)*


conv1_1

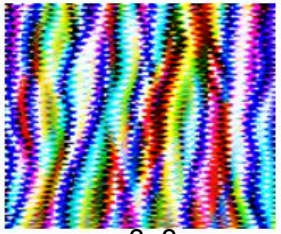

conv1_2


conv2_1


conv2_2


conv3_1


conv3_2


conv3_3


conv4_1


conv4_2

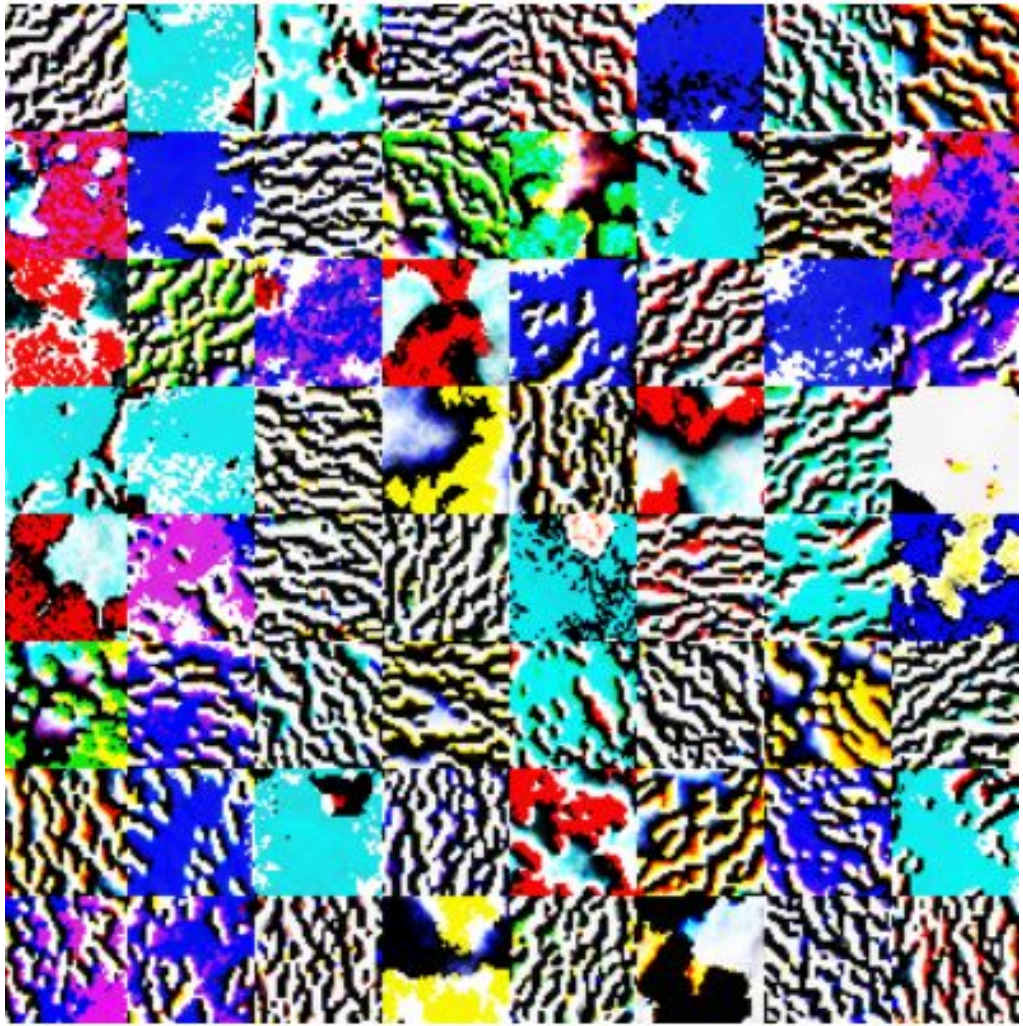

conv4_3


conv5_1

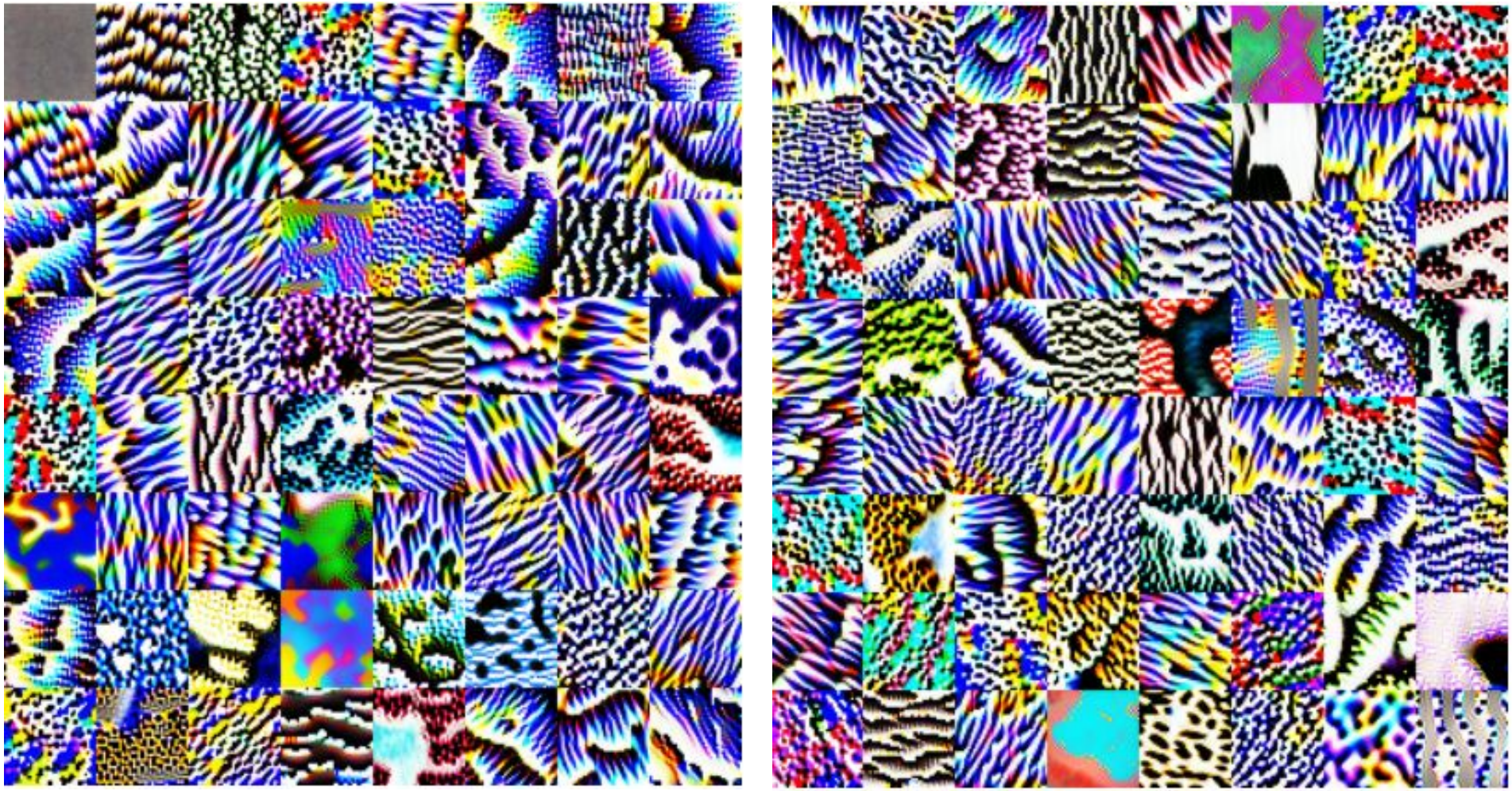

conv5_2


conv5_3

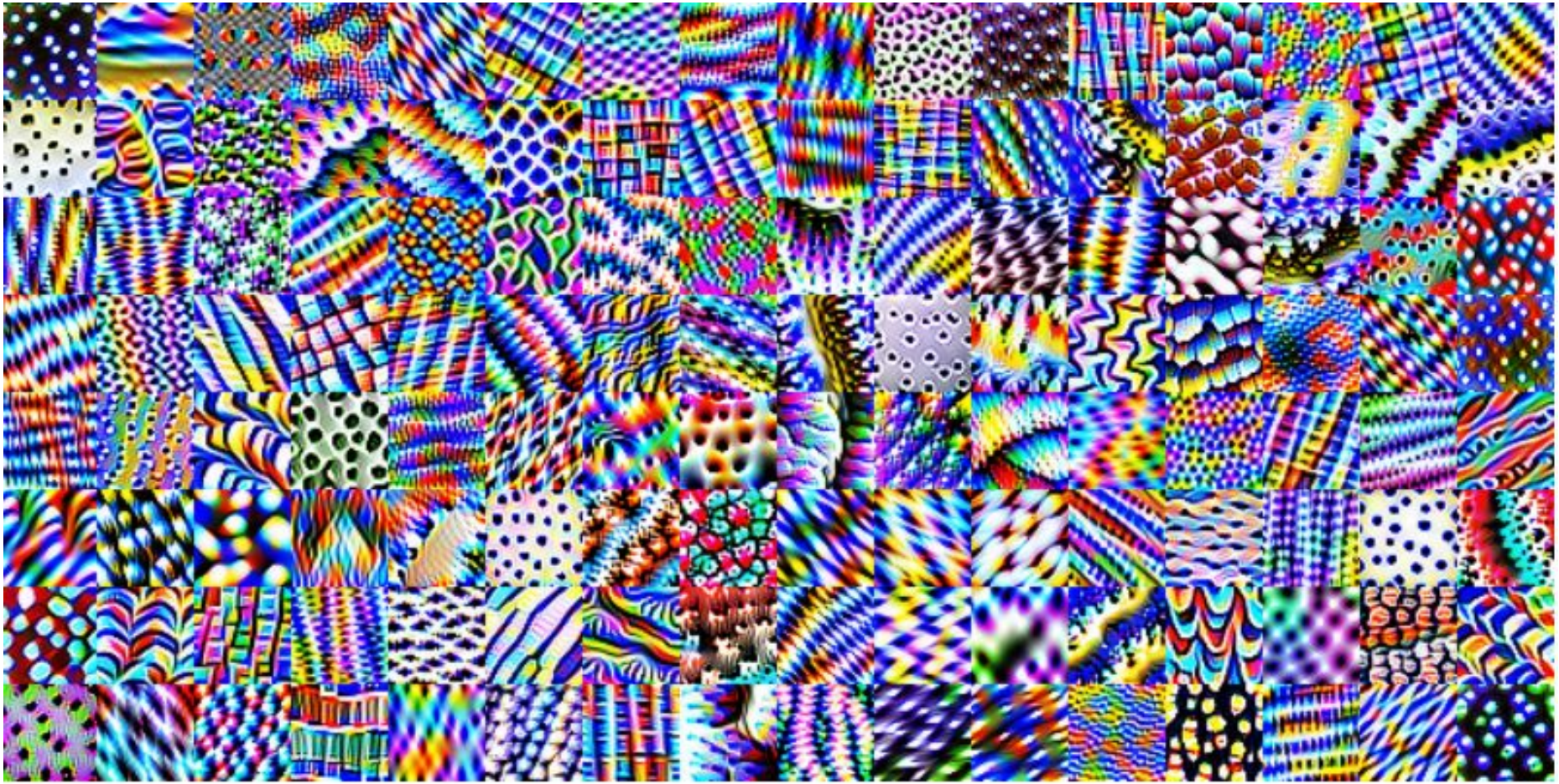
Visualisations of all neurons in layers



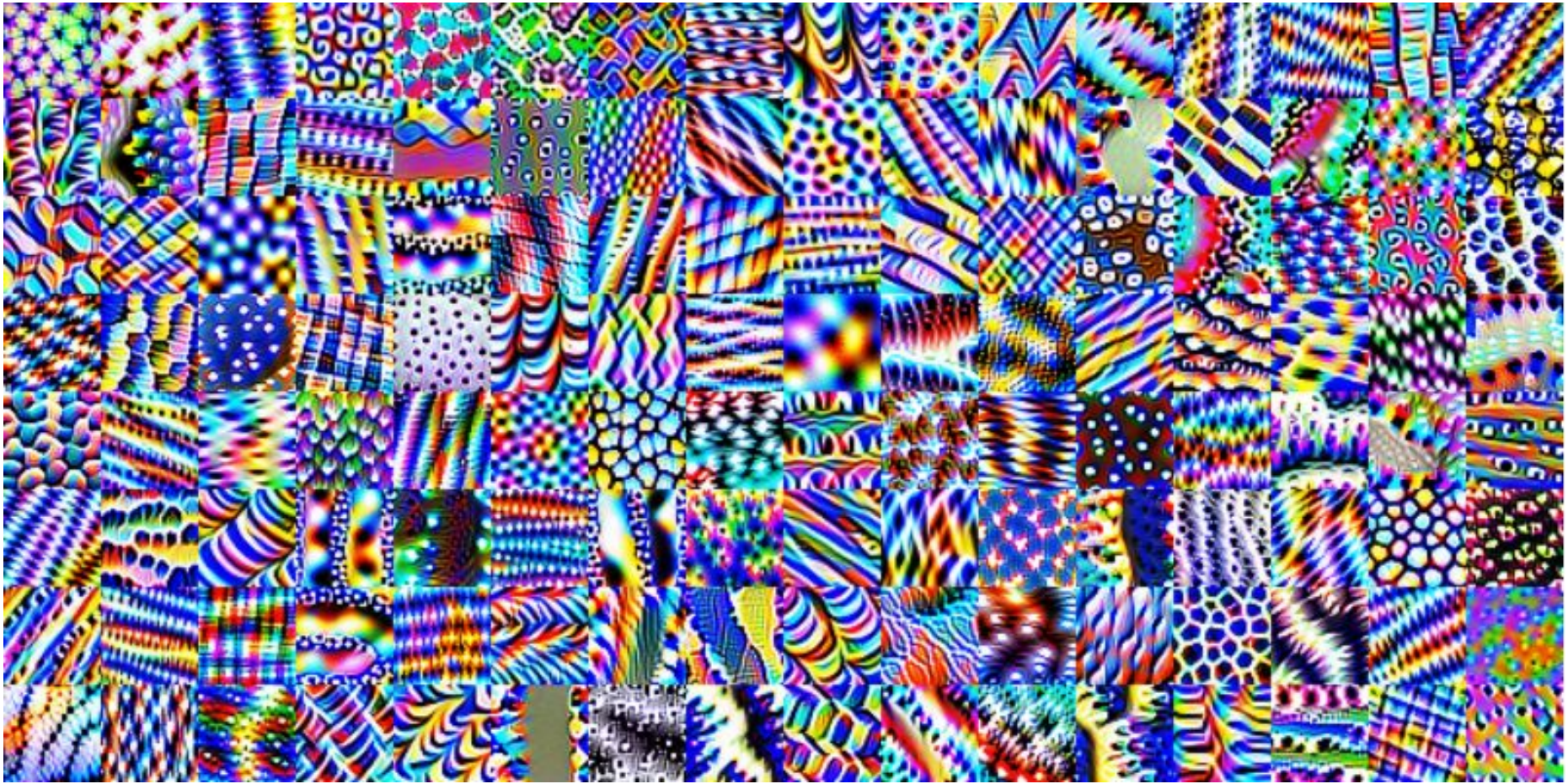
Layer 1_1



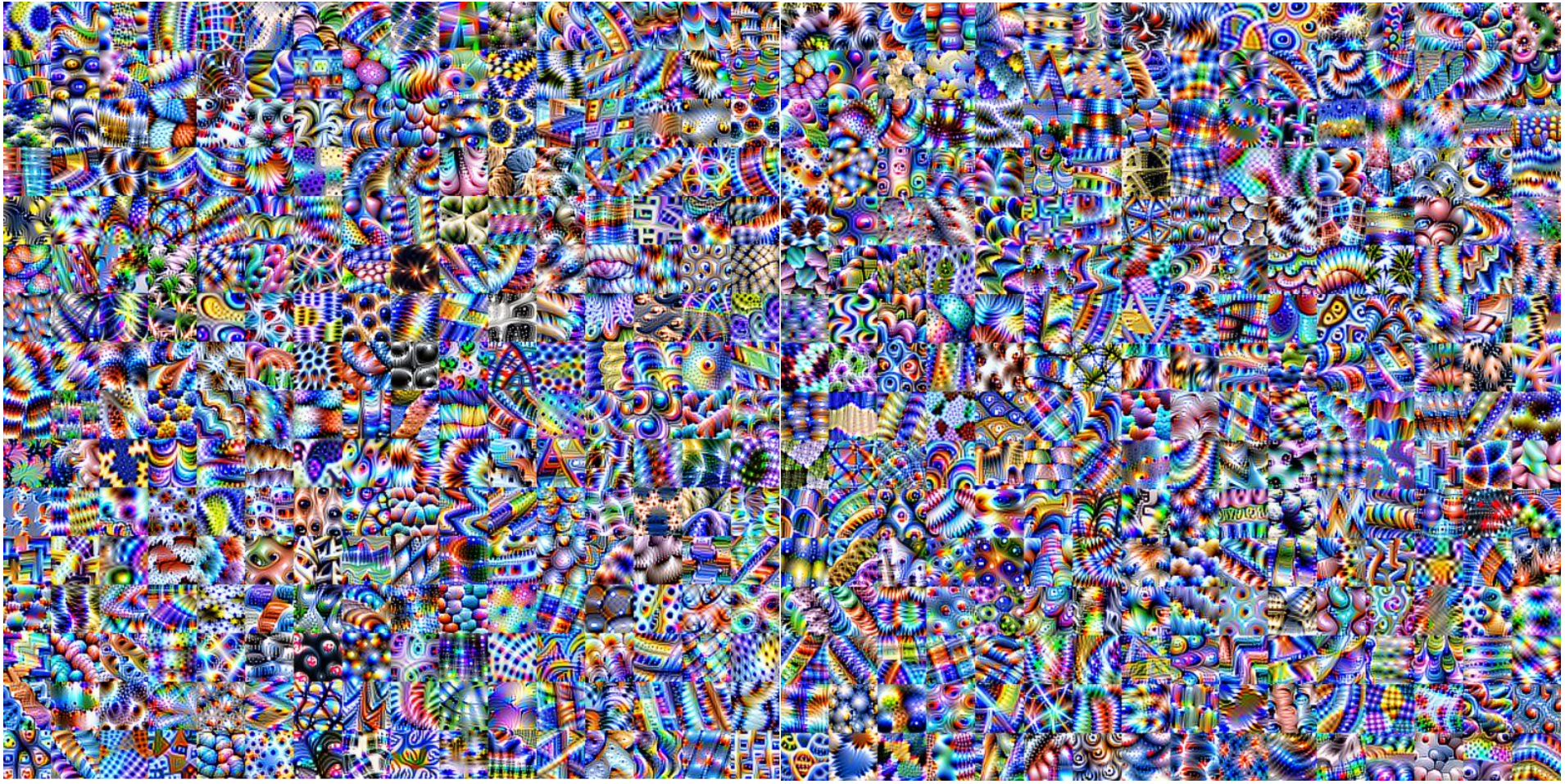
conv2_1



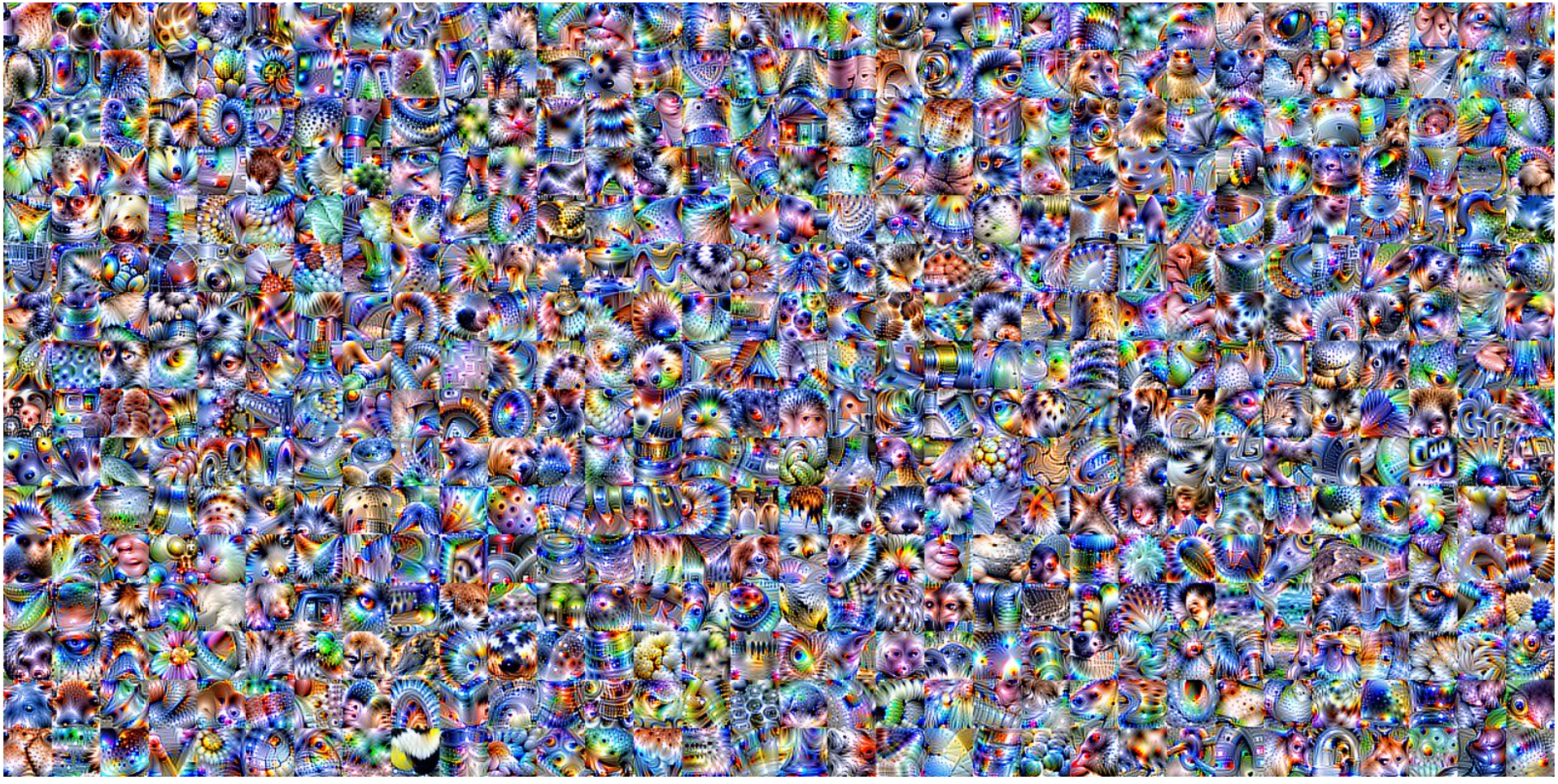
conv3_1_1



conv3_1_2



conv4_1

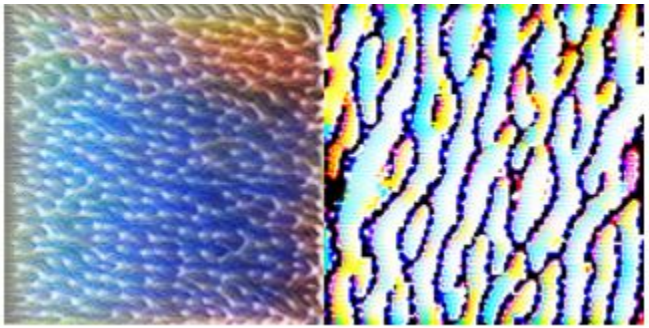


conv5_1

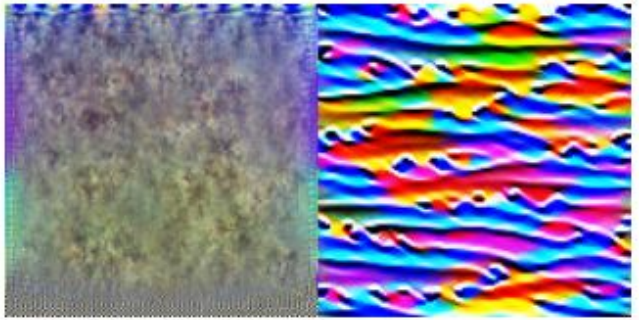
conv5_2

conv5_3

Negatively optimised neurons (Left image is the negatively optimised)

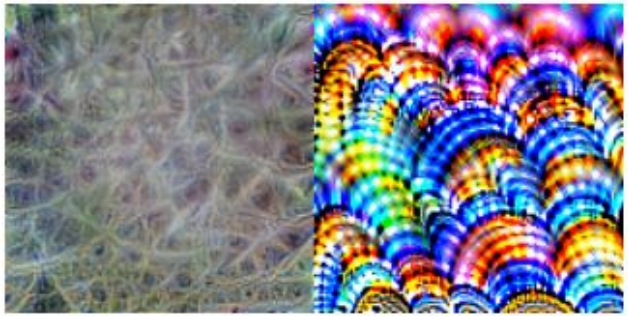


Conv2_2



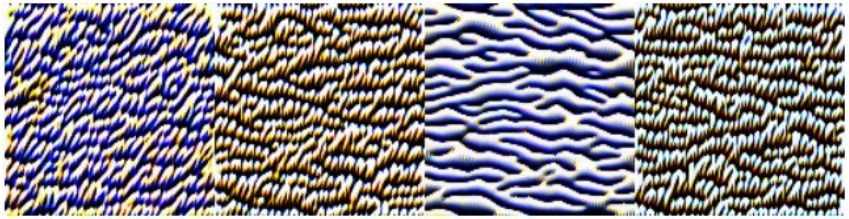
Conv3_1

Conv4_1

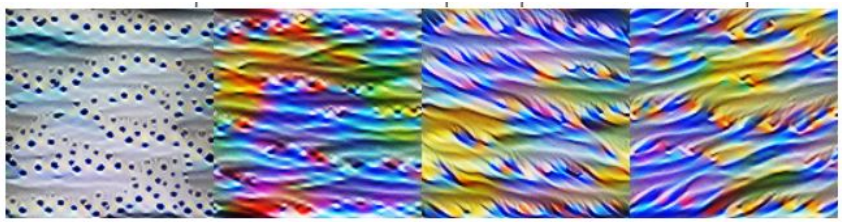


Conv4_2

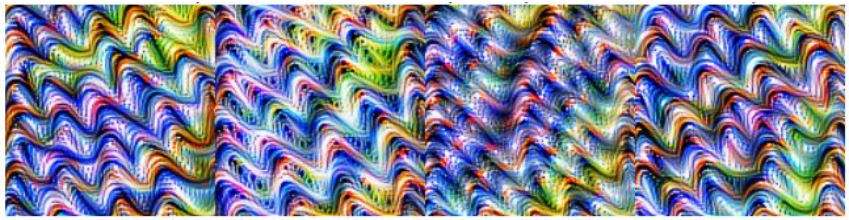
Diversity in visualisations



Conv2_1 :unit 1

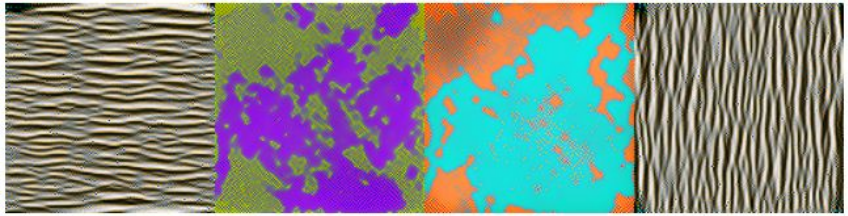


Conv3_1 :unit 1



Conv4_1 :unit 1

Conv5_1 :unit 1



Conv1_1 :unit 1

Conv5_1 :unit 4

Interaction between neurons

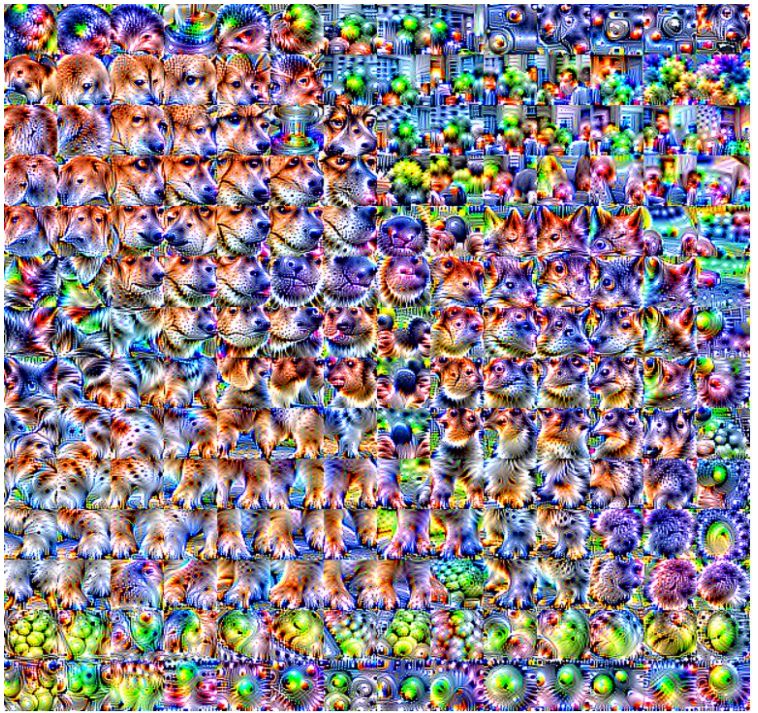
Tried interpolation between two neurons:
Neuron on the left: Layer conv5_1/conv5_1 unit 303
Neuron on the left: Layer conv5_1/conv5_1 unit 5

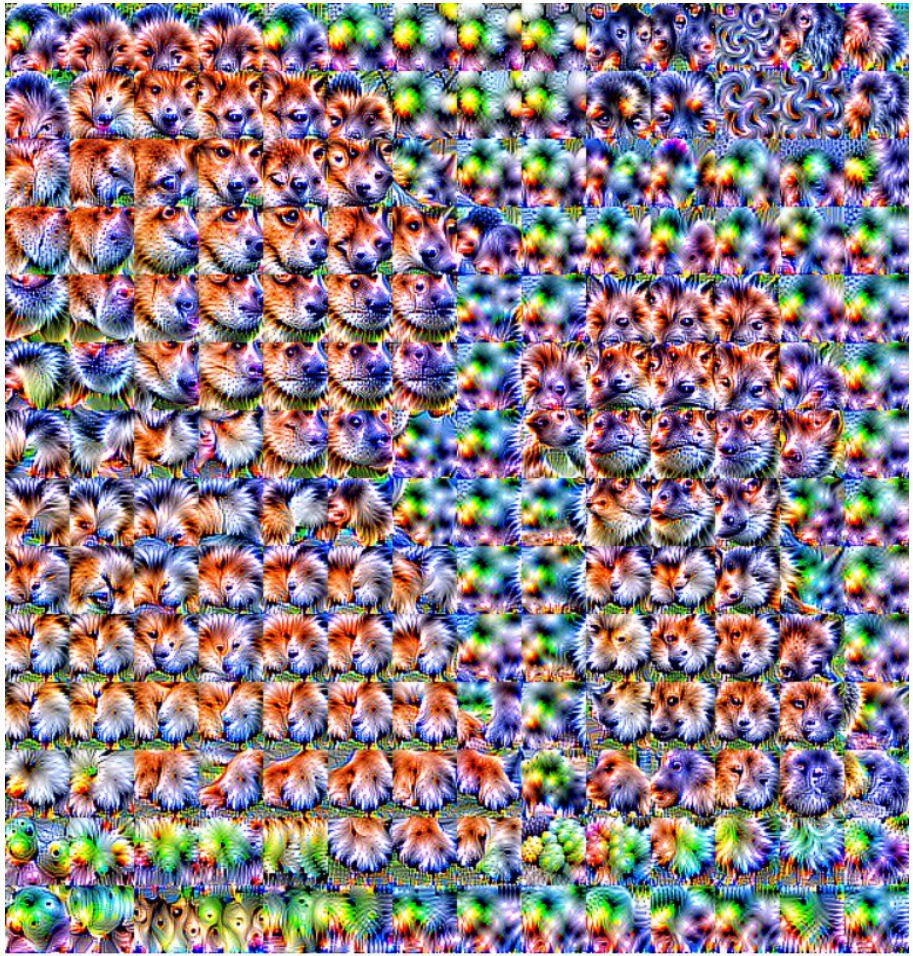
Feature Attribution

Activation grid

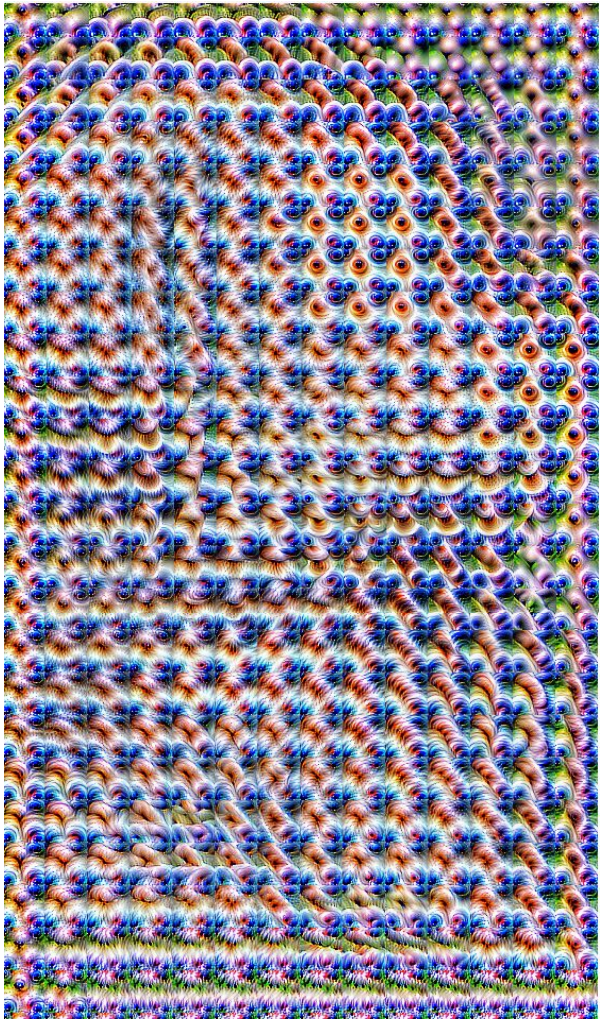
We find the activation grids at each grid point and display the visualization of the neuron giving the best visualization at that point.

Activation grid: for layer 5_2





conv5_1

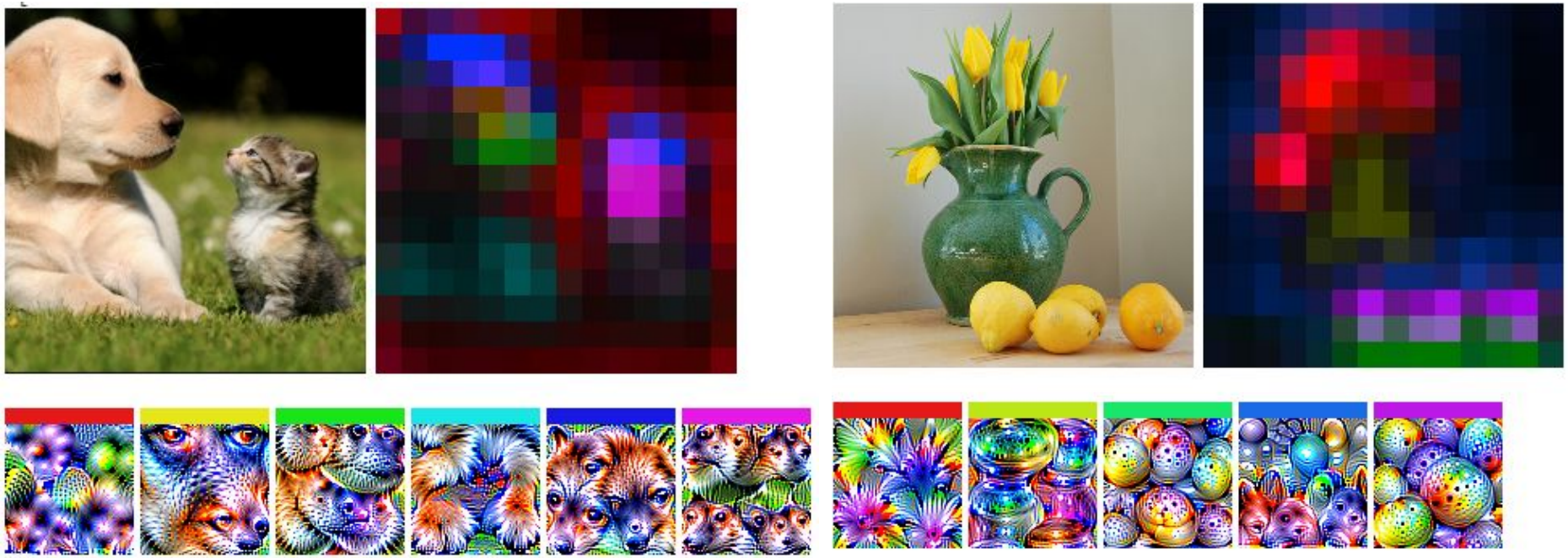


conv4_1

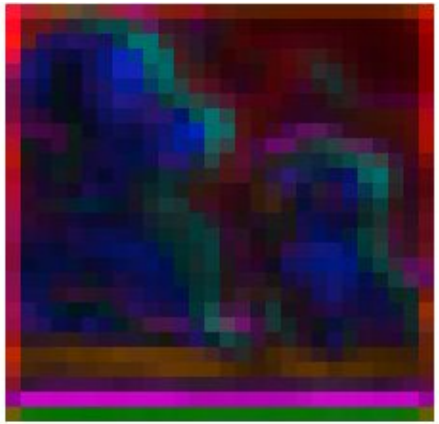
Neuron Groups

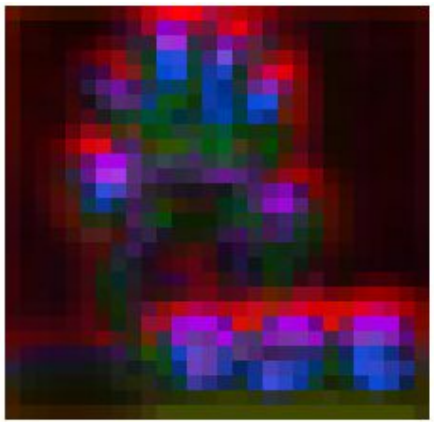
By using non-negative matrix factorization we can reduce the large number of neurons to a small set of groups that concisely summarize the story of the network.

Overwhelmingly large number of neurons can be reduced to a small set of groups, concisely summarizing the story of the neural network.

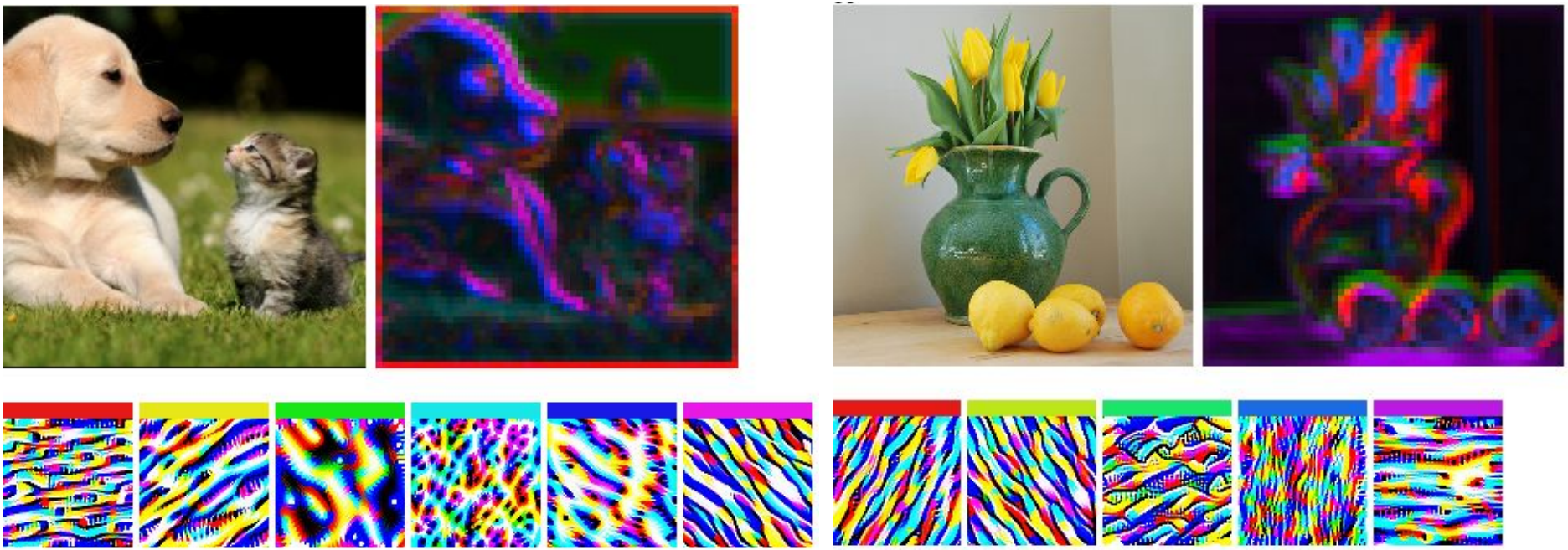


conv5_1



conv4_1



conv2_1