SUPPLEMENTAL MATERIAL: Assessing brain-like activations in convolutional neural networks

Olivia Yoo
Neuroscience Concentration, Division of Life Sciences
Harvard College
Cambridge, MA 02138
oliviayoo@college.harvard.edu

This document contains all of the relevant figures I generated for my project, which I largely did not have room to include in the original report. I include examples of all of the figures I generated for all six of the networks I investigated: AlexNet, ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152.

1 Human/CNN activity mapping

1.1 Normal images

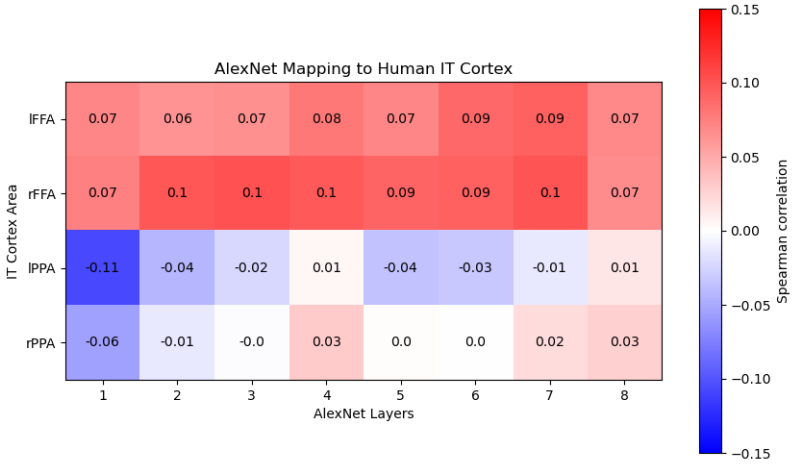

Figure 1: Mapping matrix of similarities between AlexNet layer activations and human IT cortex responses to the original 92 images

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

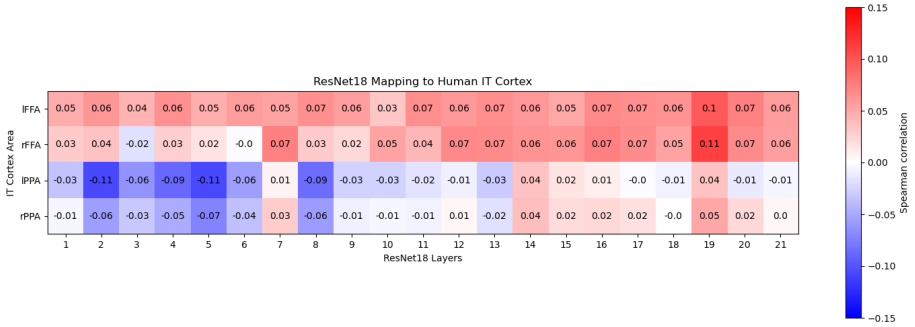

Figure 2: Mapping matrix of similarities between ResNet18 layer activations and human IT cortex responses to the original 92 images

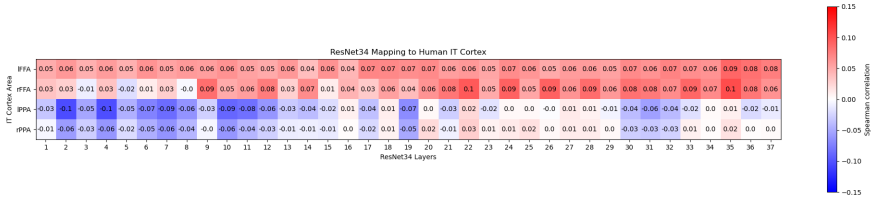

Figure 3: Mapping matrix of similarities between ResNet34 layer activations and human IT cortex responses to the original 92 images

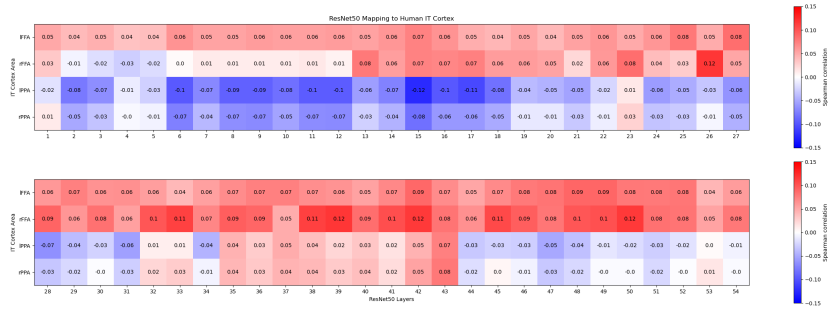

Figure 4: Mapping matrix of similarities between ResNet50 layer activations and human IT cortex responses to the original 92 images

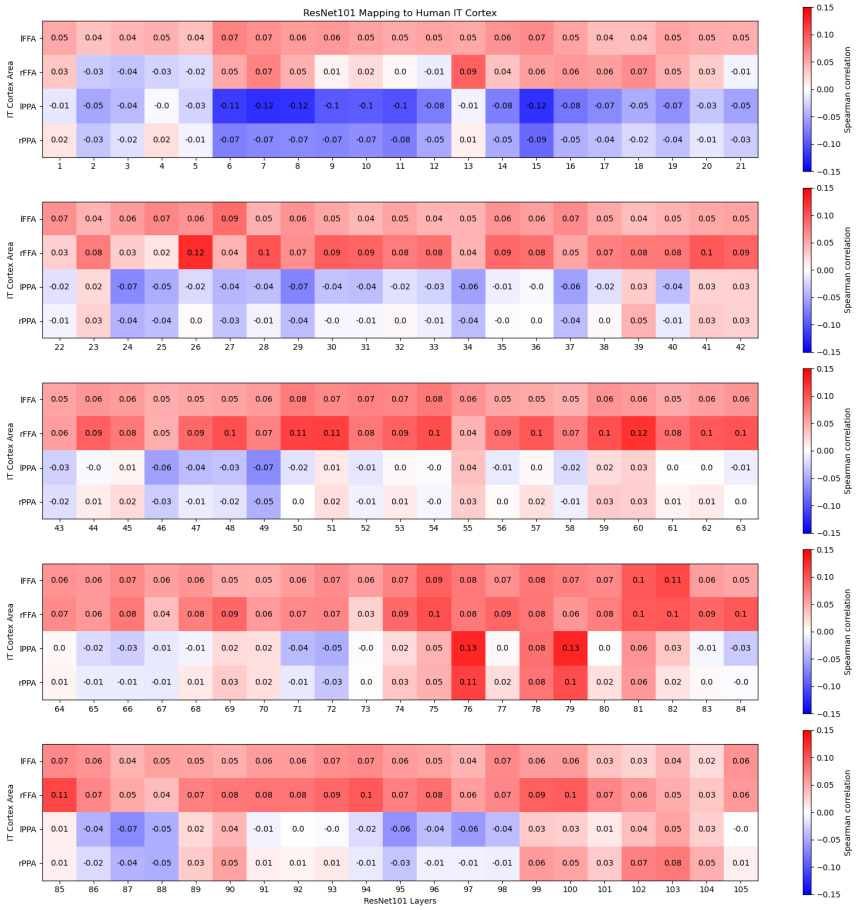

Figure 5: Mapping matrix of similarities between ResNet101 layer activations and human IT cortex responses to the original 92 images

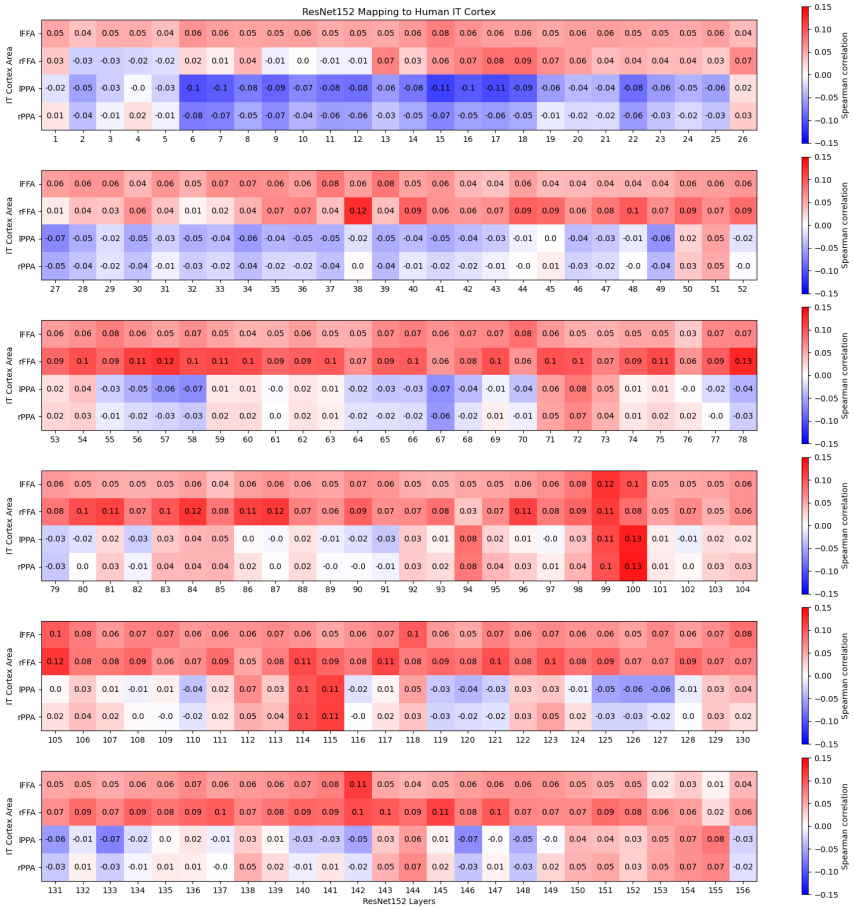

Figure 6: Mapping matrix of similarities between ResNet152 layer activations and human IT cortex responses to the original 92 images

1.2 Modified images

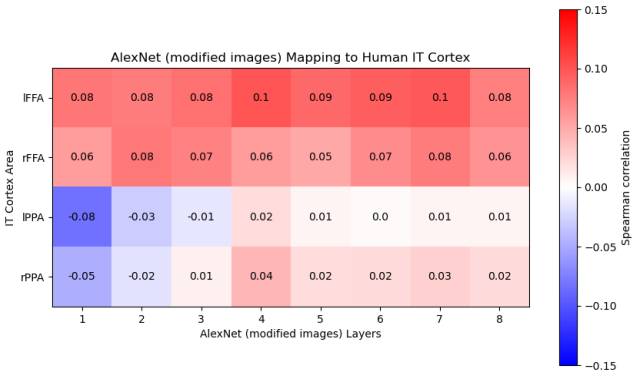

Figure 7: Mapping matrix of similarities between AlexNet layer activations and human IT cortex responses to the modified 92 images

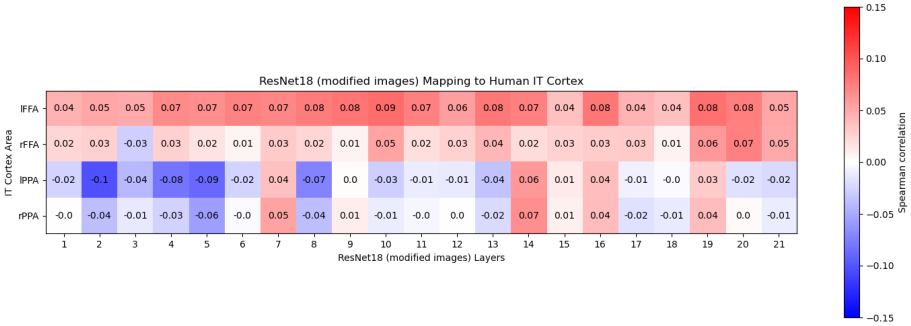

Figure 8: Mapping matrix of similarities between ResNet18 layer activations and human IT cortex responses to the modified 92 images

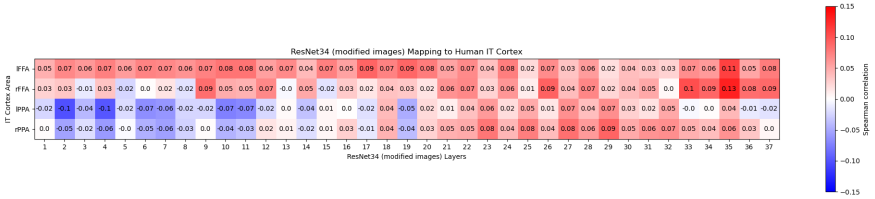

Figure 9: Mapping matrix of similarities between ResNet34 layer activations and human IT cortex responses to the modified 92 images

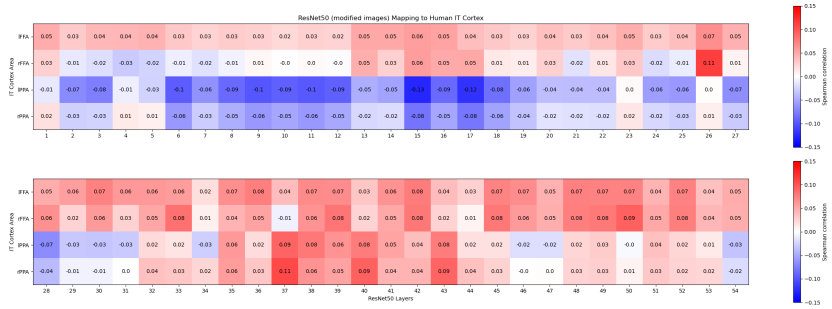

Figure 10: Mapping matrix of similarities between ResNet50 layer activations and human IT cortex responses to the modified 92 images

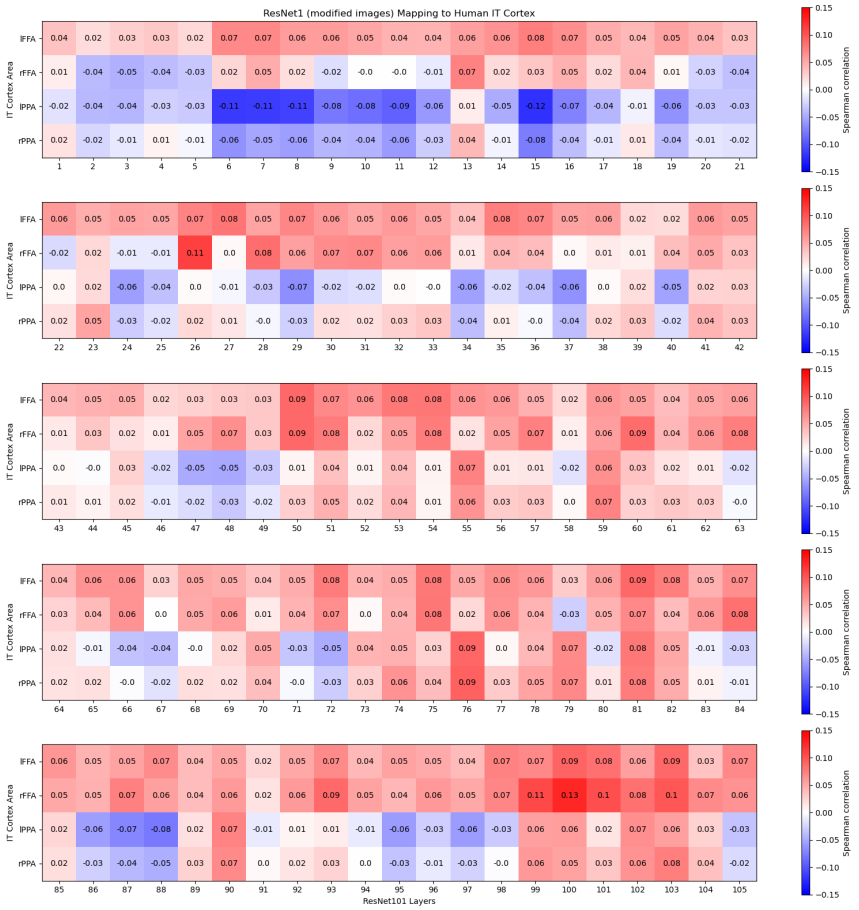

Figure 11: Mapping matrix of similarities between ResNet101 layer activations and human IT cortex responses to the modified 92 images

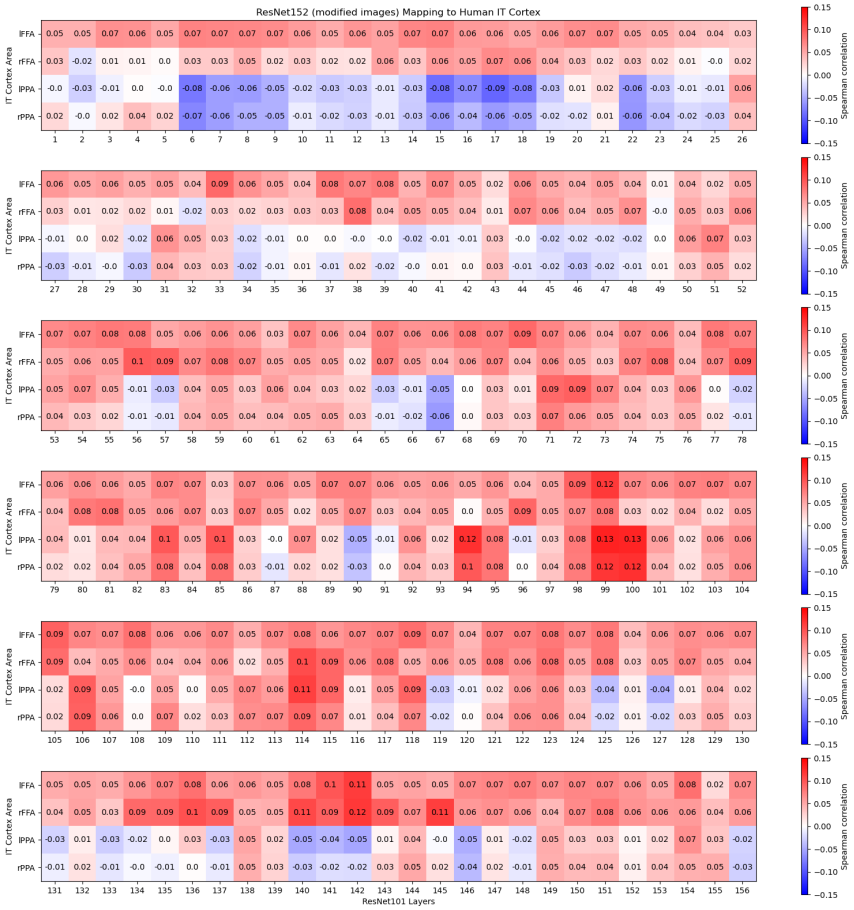

Figure 12: Mapping matrix of similarities between ResNet152 layer activations and human IT cortex responses to the modified 92 images

2 Network layer similarities

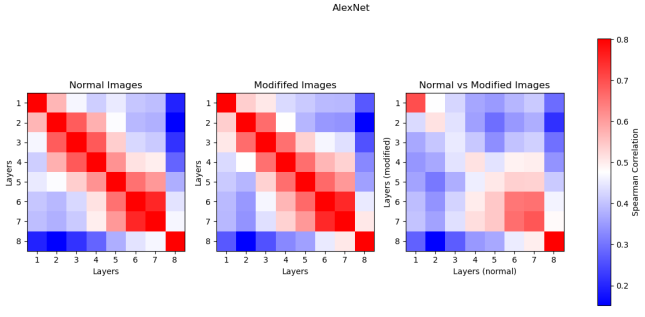

Figure 13: Similarities between the RDMs of layer activations for AlexNet in response to both the normal and modified natural images. Left: a similarity matrix between the normal responses to itself. Middle: a similarity matrix between the modified responses to itself. Right: a similarity matrix comparing the responses to the normal and modified images

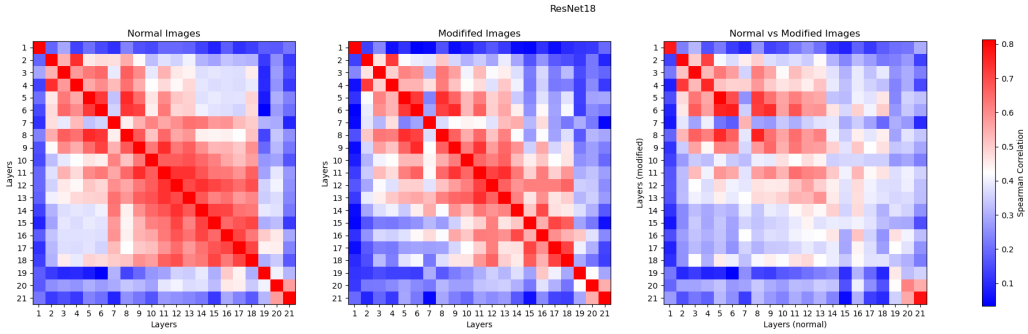

Figure 14: Similarities between the RDMs of layer activations for ResNet18 in response to both the normal and modified natural images. Left: a similarity matrix between the normal responses to itself. Middle: a similarity matrix between the modified responses to itself. Right: a similarity matrix comparing the responses to the normal and modified images

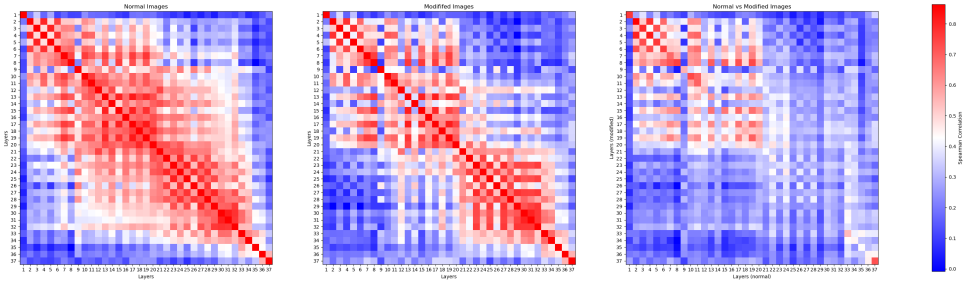

Figure 15: Similarities between the RDMs of layer activations for ResNet34 in response to both the normal and modified natural images. Left: a similarity matrix between the normal responses to itself. Middle: a similarity matrix between the modified responses to itself. Right: a similarity matrix comparing the responses to the normal and modified images

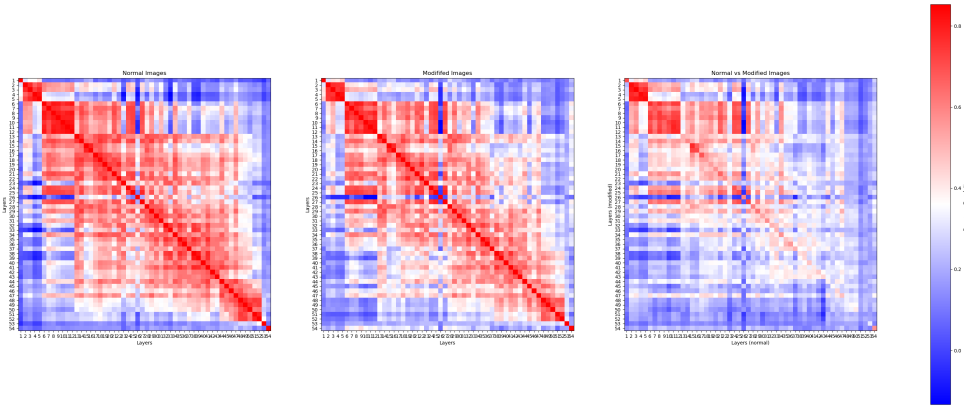

Figure 16: Similarities between the RDMs of layer activations for ResNet50 in response to both the normal and modified natural images. Left: a similarity matrix between the normal responses to itself. Middle: a similarity matrix between the modified responses to itself. Right: a similarity matrix comparing the responses to the normal and modified images

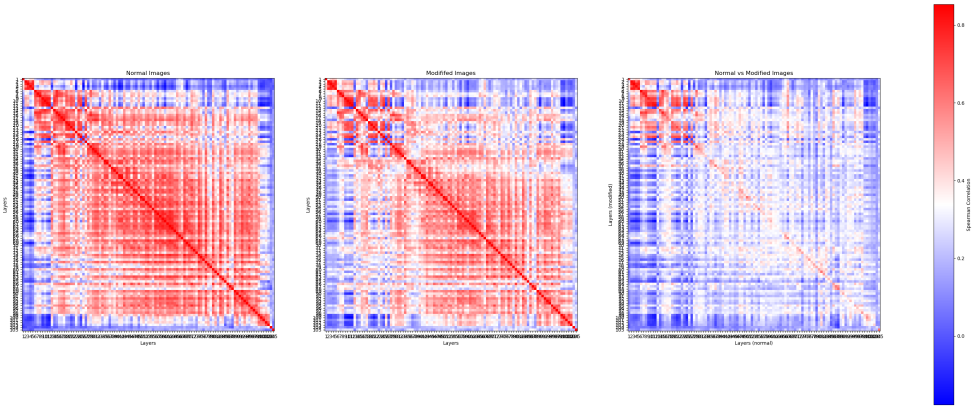

Figure 17: Similarities between the RDMs of layer activations for ResNet101 in response to both the normal and modified natural images. Left: a similarity matrix between the normal responses to itself. Middle: a similarity matrix between the modified responses to itself. Right: a similarity matrix comparing the responses to the normal and modified images

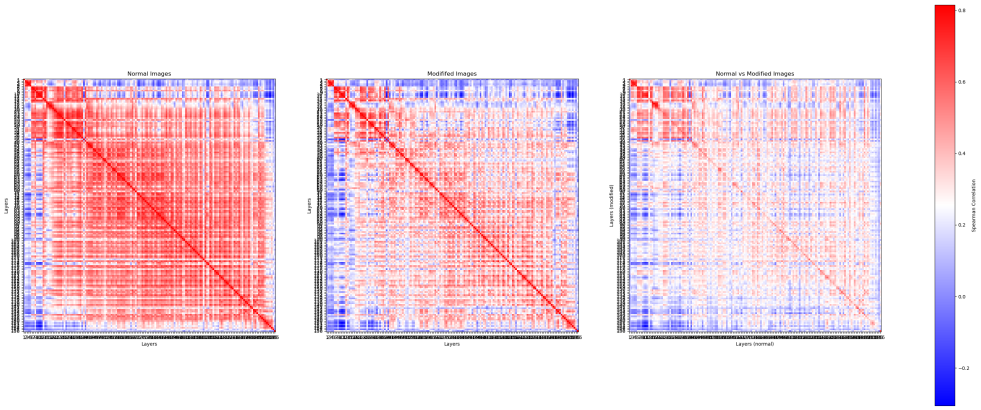

Figure 18: Similarities between the RDMs of layer activations for ResNet152 in response to both the normal and modified natural images. Left: a similarity matrix between the normal responses to itself. Middle: a similarity matrix between the modified responses to itself. Right: a similarity matrix comparing the responses to the normal and modified images