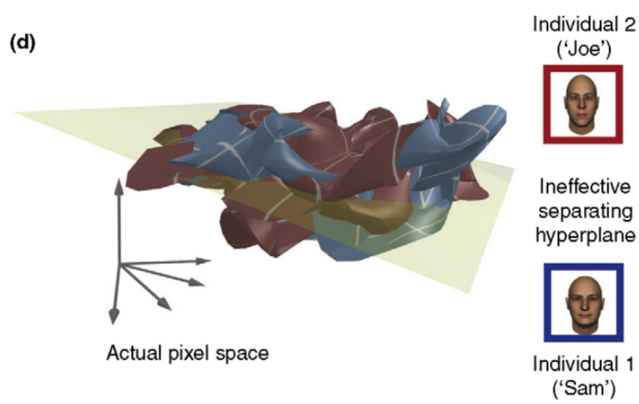
<u>Capstone Project - Manifold Estimation in CNNs during Object Recognition</u>

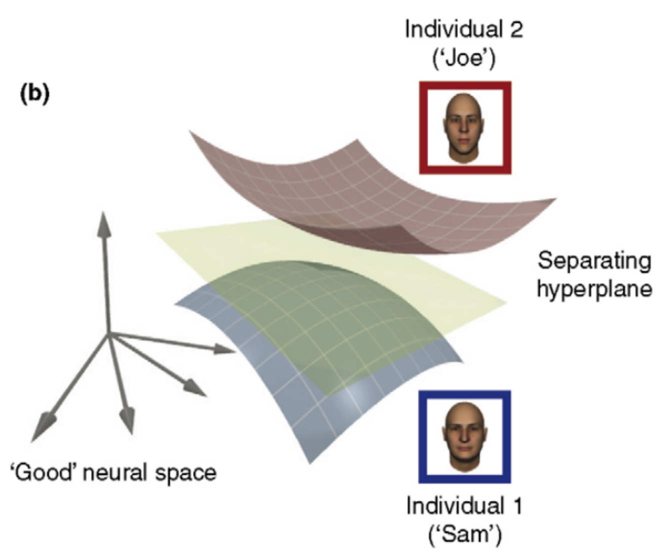
Background

Theoretical Neuroscience literature suggests a notion of Neural Manifolds that are task and behviour specific. The idea is that when the primate brain is presented a visual stimulus, the neurons in the visual pathway fire a certain way. If you record the neural activity of each neuron in the ventral visual stream, you get a point in an n-dimensional space where n is the number of neurons recorded. Literature suggests that the visual representation of an object (despite any identity-preserving variances), exists as a lower-dimensional manifold within this high dimensional space.

Other work suggests that the ventral visual pathway in the primate brain is tasked with untangling these manifolds. The image below shows a rendition of 'tangled' manifolds of two distinct people generated by capturing images of the subject in varying settings / perspectives.



In order to easily distinguish between the two subjects, the neural space must be 'transformed' in a way that the manifolds are easily separable by a linear decision function (a hyperplane in the space).



The theory suggests that the goal of the visual pathway (in terms of object recognition) is to transform the representation space from something that is easy to record (like the RGB channels of an image or the activations of a retinal cell) to something that allows for easy distinction between, and consequently recognition of, objects.

Project Implementation

In my Capstone Project, I want to verify this phenomenon in CNNs that are trained to effectively and accurately distinguish between two images. I hope to do this by training a CNN on classes of objects with certain meaningful variance in the images. This will be followed by subsequent recording of each neural unit in the CNN during a forward pass of an image. Each forward pass of an object will result in an m-dimensional vector where m is the number of neural units measured. Capturing these vectors for numerous images of the same object will result in a manifold using techniques like Isomap and Locally Linear Embeddings (LLE).

Through the project, I want to deliver the following:

- A visual representation of the manifolds of two (or more) object-invariant manifolds
    - Visualization by dimensionality-reduction techniques may not lead to a meaningful result
    - An alternate approach could be using a distance-based metric to verify the 'difference' in the manifolds
- Investigate how the manifolds of these objects change over the course of the layers of the CNN. Do they become more separable the deeper you go?
    - The verification of this result depends on the previous deliverable of a representation of the manifolds

References

1. Untangling invariant object recognition
2. How does the brain solve visual object recognition?
3. Separability and Geometry of Object Manifolds in Deep Neural Networks
4. Neural Constraints on Learning