Brain-like Object Manifold Separation in Deep Neural Networks

Capstone Project Presentation by **Niranjan Rajesh**
Under the Supervision of **Professors Debayan Gupta and Venkat Ramaswamy**

Table of contents

01

Motivation

Setting the scene – **solving visual intelligence**

Solving Visual Intelligence - Bridging the gap between Machine and Minds

- Draw **inspiration** from Neuroscience to advance AI?
- Can we **adapt** learnings from how the brain solves object recognition to computer vision?

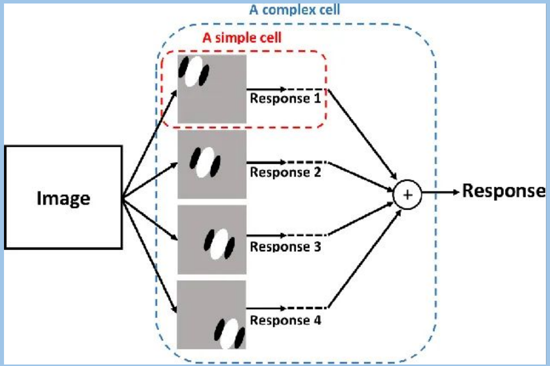
Computer Vision Today – Very Deep Networks

- State-of-the-art today:
  - **Convolutional Neural Networks (CNNs)**
  - Vision Transformers

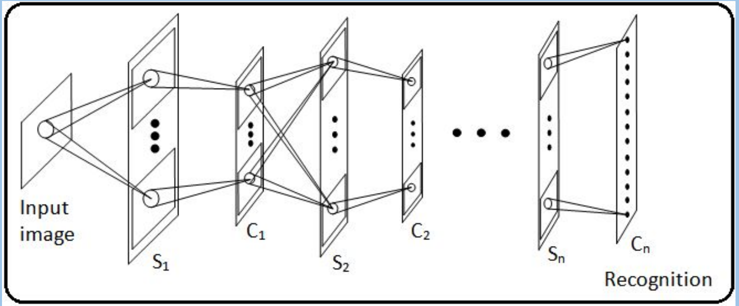
- Object Recognition: **Image Classification** and Semantic Segmentation

CNNs – Historically Inspired by Neuroscience

- Hubel and Wiesel (1959) - Simple and Complex cells
- Fukushima (1980) - the Neocognitron
- LeCun (1989) - LeNet, the first ‘convolutional’ neural network
- AlexNet, ResNets, VGGs, MobileNets, …

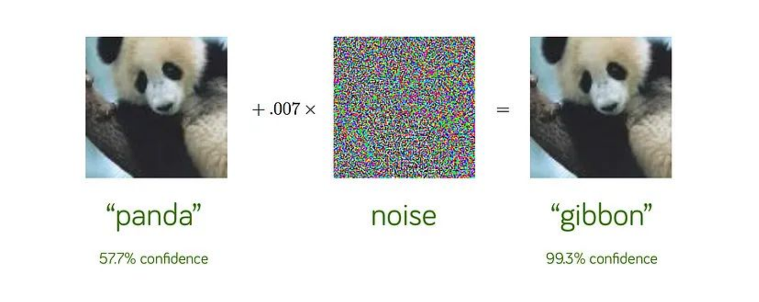


Hubel, D. H., & Wiesel, T. N. (1959)



Fukushima, K. (1980)

The Gap: Where CNNs fall short



Bridging the Gap by turning back to the Brain

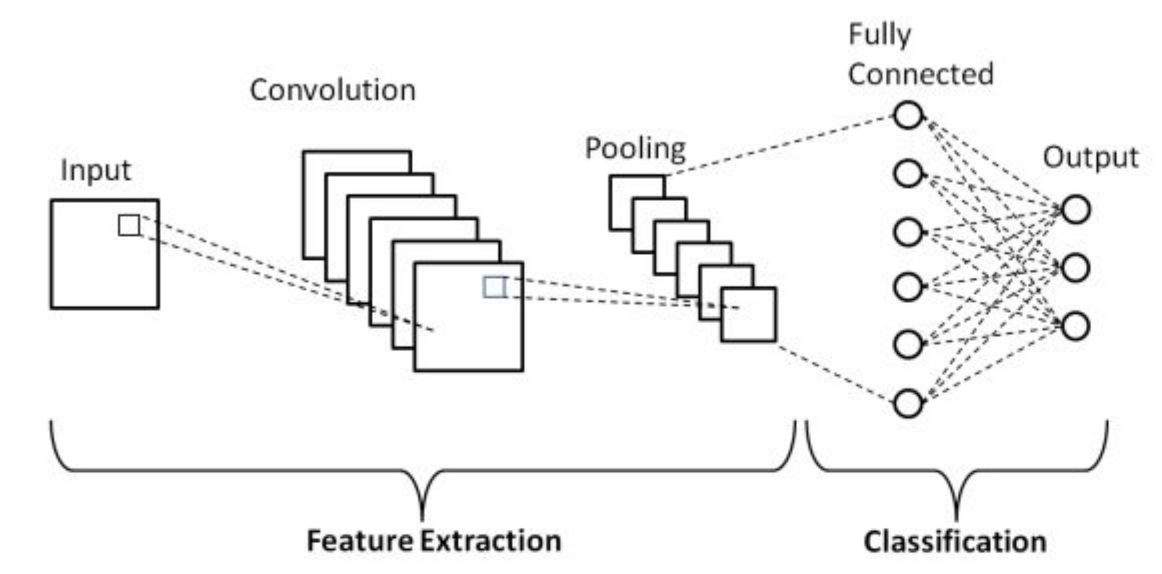
- Understand the **visual information processing** of the primate visual system to build better networks
- **Neural Manifolds** – a very helpful tool in uncovering the secrets of stimuli representations in neural networks

02

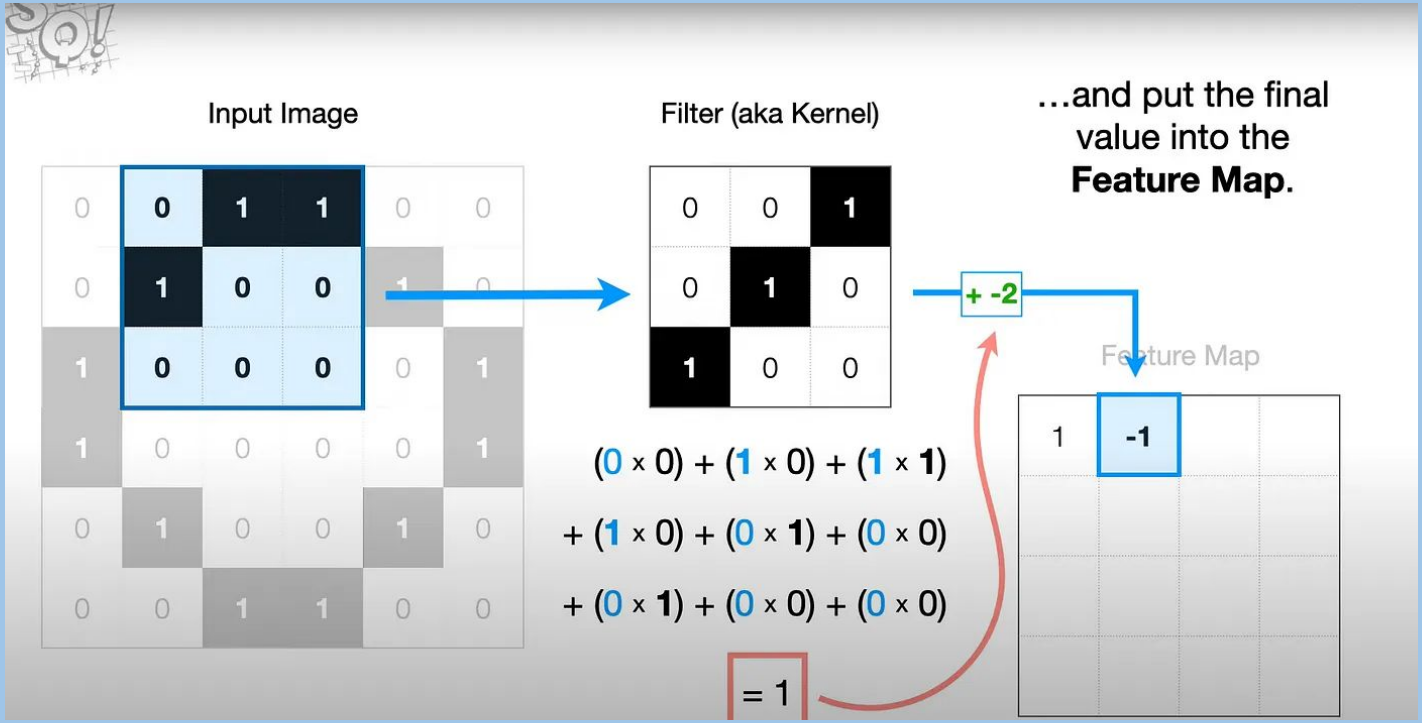
Technical Background

CNNs, Primate Visual Pathways, Intrinsic Manifolds

A Quick CNN Crash Course

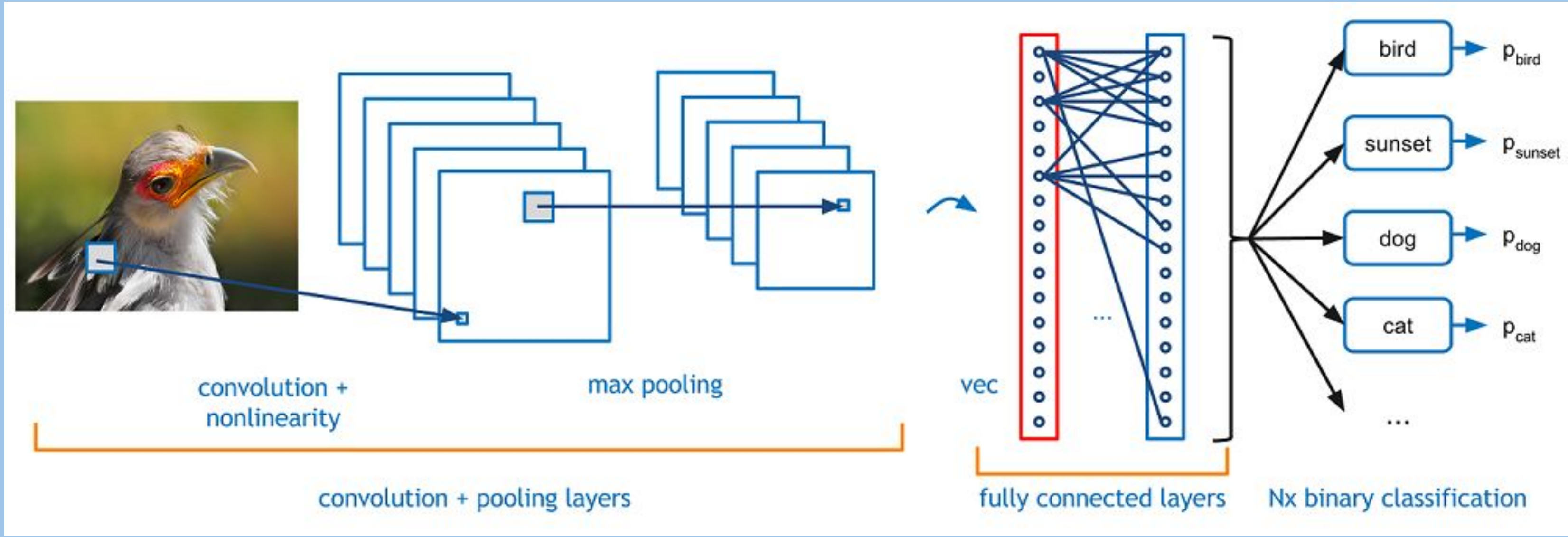


CNN Crash Course - Convolutions

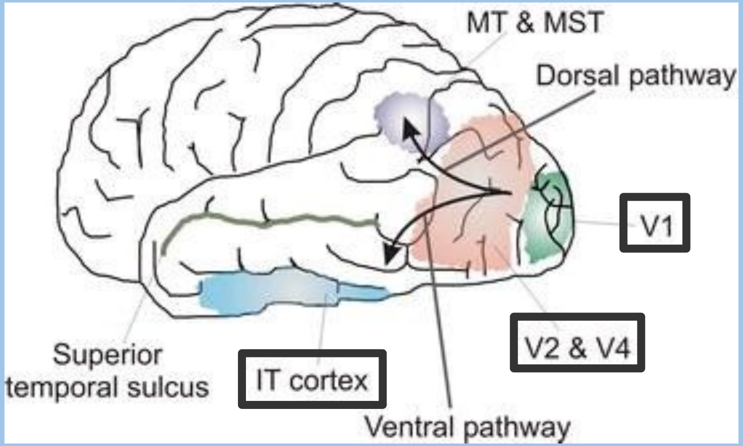


(StatQuest)

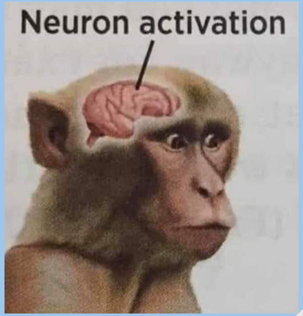
CNN Crash Course



Primate Visual Pathways

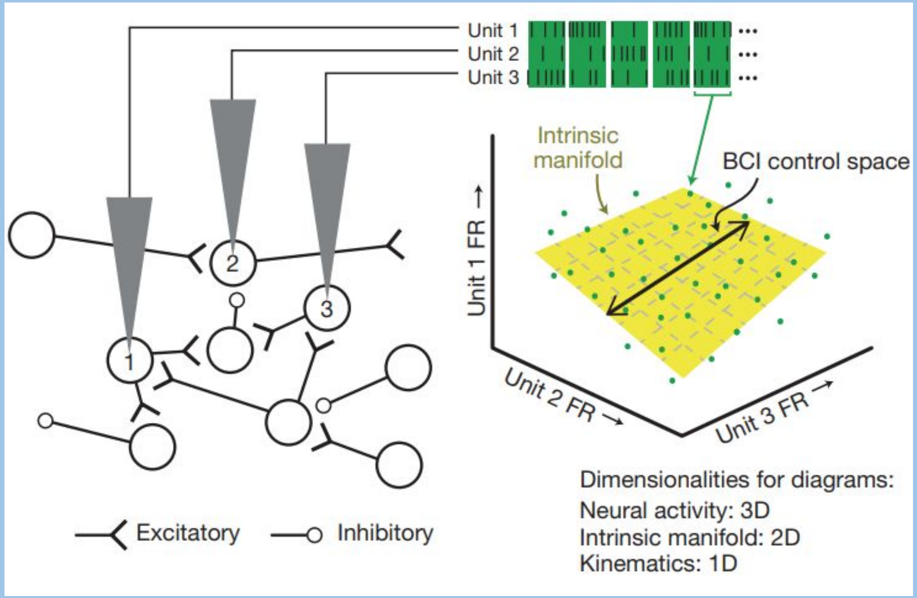


V1 → V2 → V4 → IT



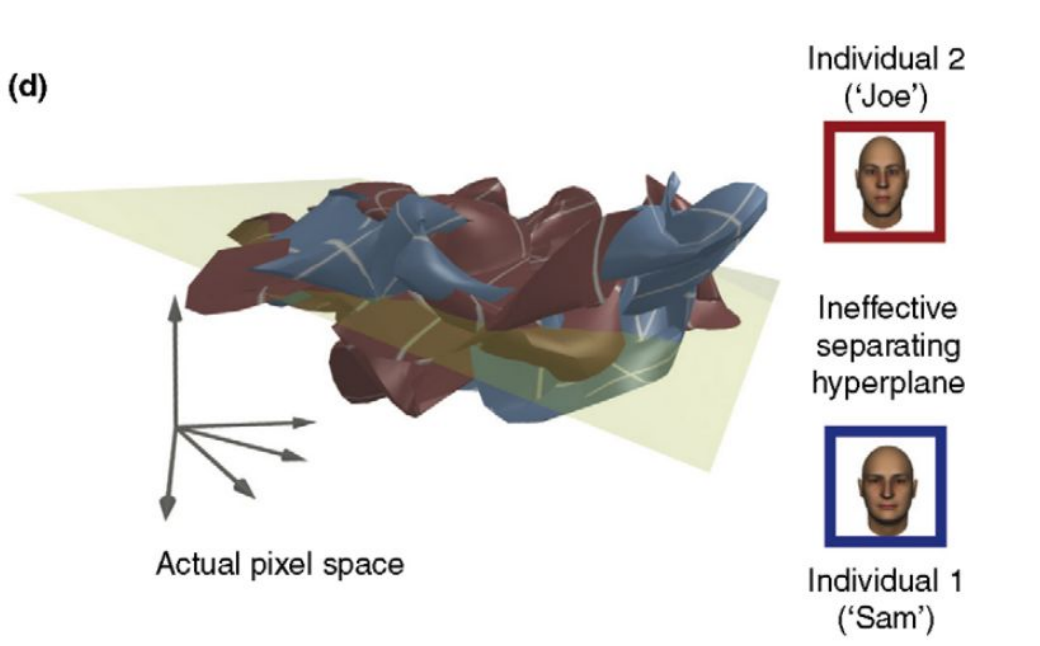
Neural Manifolds

- Task-specific subspaces in the larger neural space



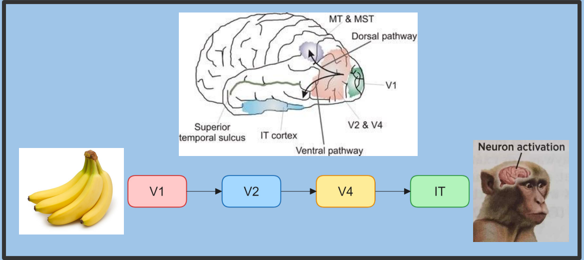
(Sadtler et al, 2014)

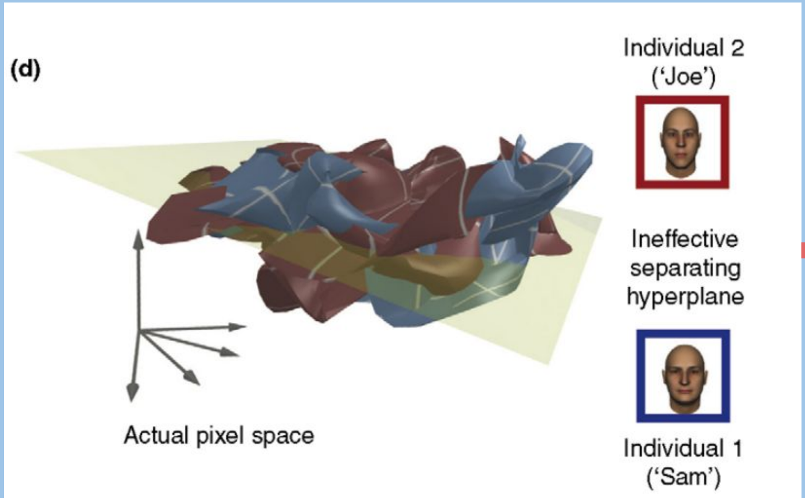
Manifolds in Vision

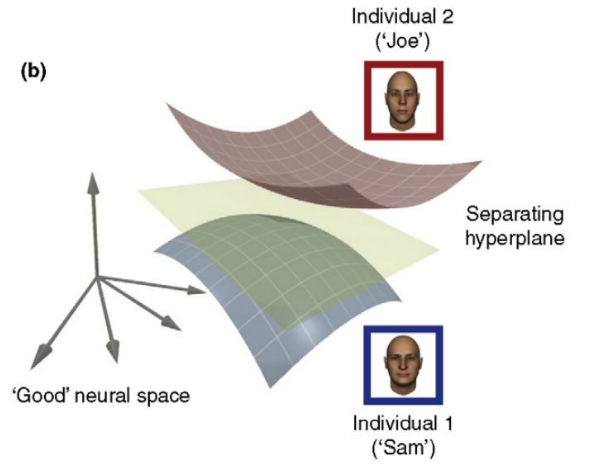


DiCarlo, J. J., & Cox, D. D. (2007)

The Goal of the Visual System - Disentanglement



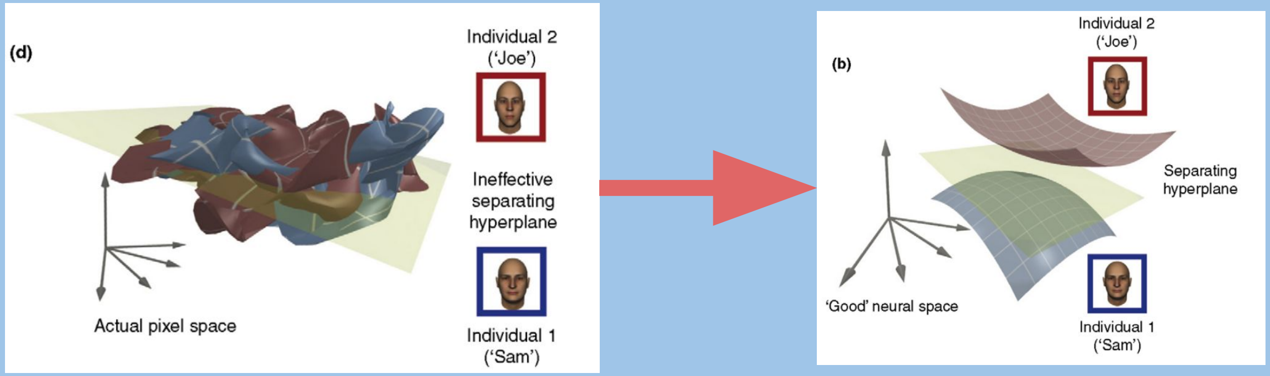




[DiCarlo, J. J., & Cox, D. D. (2007)]()

The Goal of the Primate Visual System

- Transform representations from ‘difficult to decode’ to ‘easy to decode’
- Separation of Object Manifolds over regions in the visual pathway



My Research Question

- Bridging the gap — comparing visual information processing

“Does a CNN’s layer-wise Object Manifolds get more linearly separable the deeper you go?”

Title Review: Brain-like Object Manifold Separation in Deep Neural Networks

03

Methodology

Experiment Design, Obstacles and Solutions

Answering the question

"Does a CNN's layer-wise object manifolds get more linearly separable the deeper you go?"

Algorithm:

1. Train a CNN on an Image Classification dataset
2. Get layer-wise activations (manifold points) for two different objects
3. Verify degree of linear separability for each layer activations

First Attempt

Initial Solution:

- Custom Object-Invariant Dataset of two objects
- Obtain a pretrained-CNN to classify these objects
- Run my algorithm on it

Problem: ***Always linearly separable!*** Why?

The “Blessing” of Dimensionality

8

Trivially, if you have N data points, they will be linearly separable in $N-1$ dimensions. Any structure in the data may reduce the required dimensionality for linear separation further.

Share Cite Improve this answer Follow

edited Jul 31, 2012 at 23:37

answered Jul 31, 2012 at 22:18

MLS

738 3 13

Stats Stack Exchange

6000 data points and 13,000 dimensions

New Approach

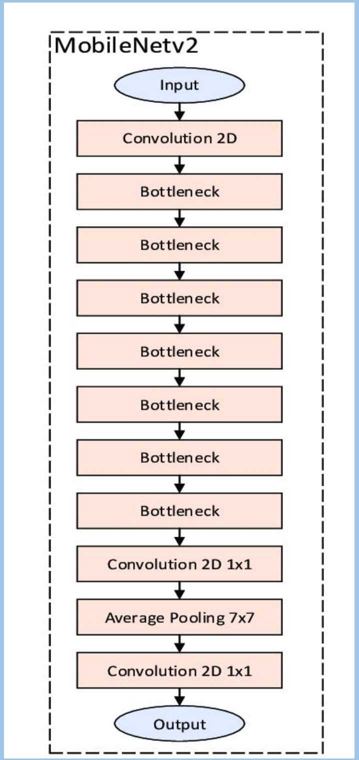
> ***"Does a CNN's layer-wise class manifolds get more linearly separable the deeper you go?"***

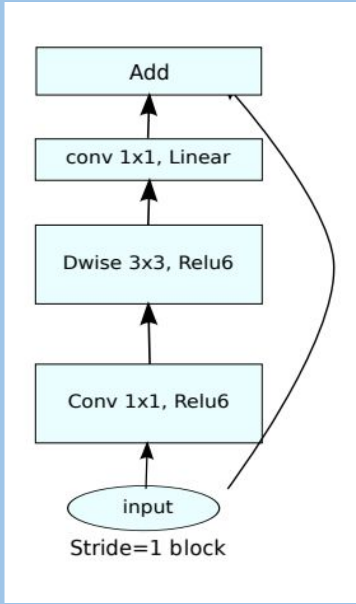
Algorithm:

1. Obtain a CNN and dataset keeping in mind the Dimension problem
2. Train the CNN on the Image Classification dataset
3. Get layer-wise activations (manifold points) for two different objects
4. Verify degree of linear separability for each layer activations

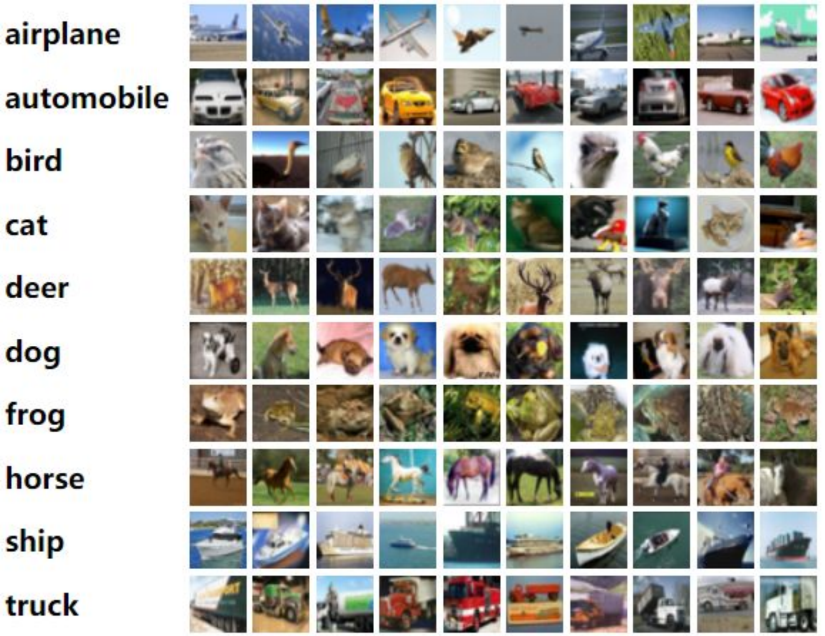
Architecture and Dataset

10,000 data points and <8k dimensions



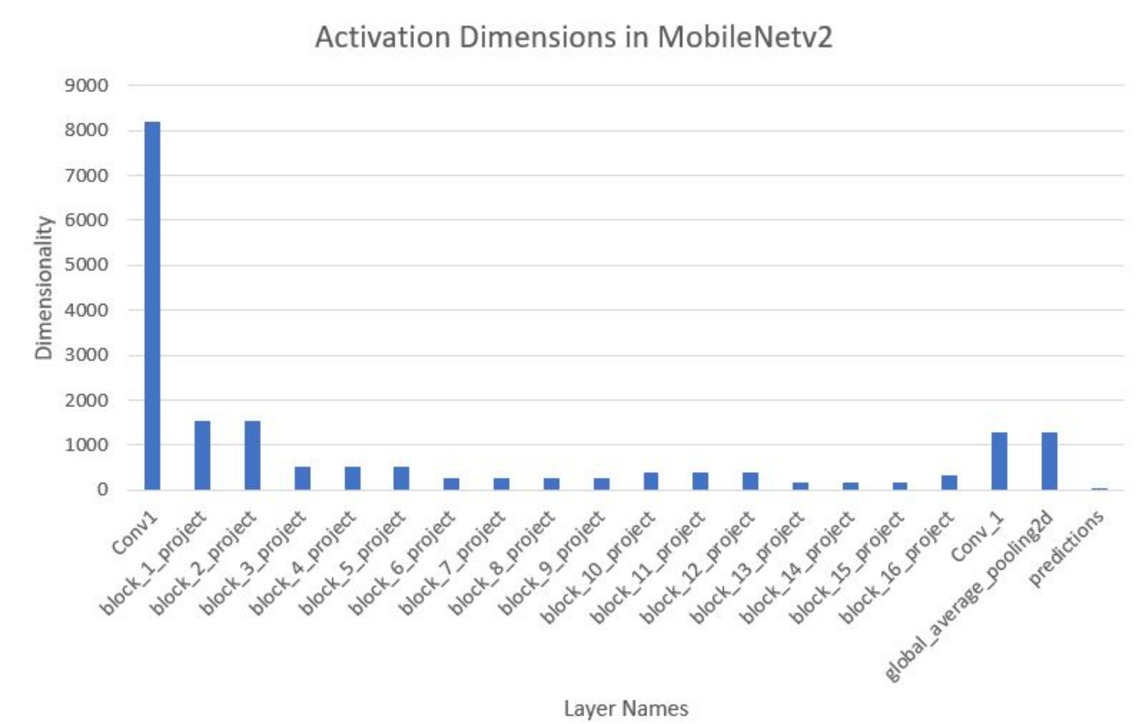


MobileNet v2



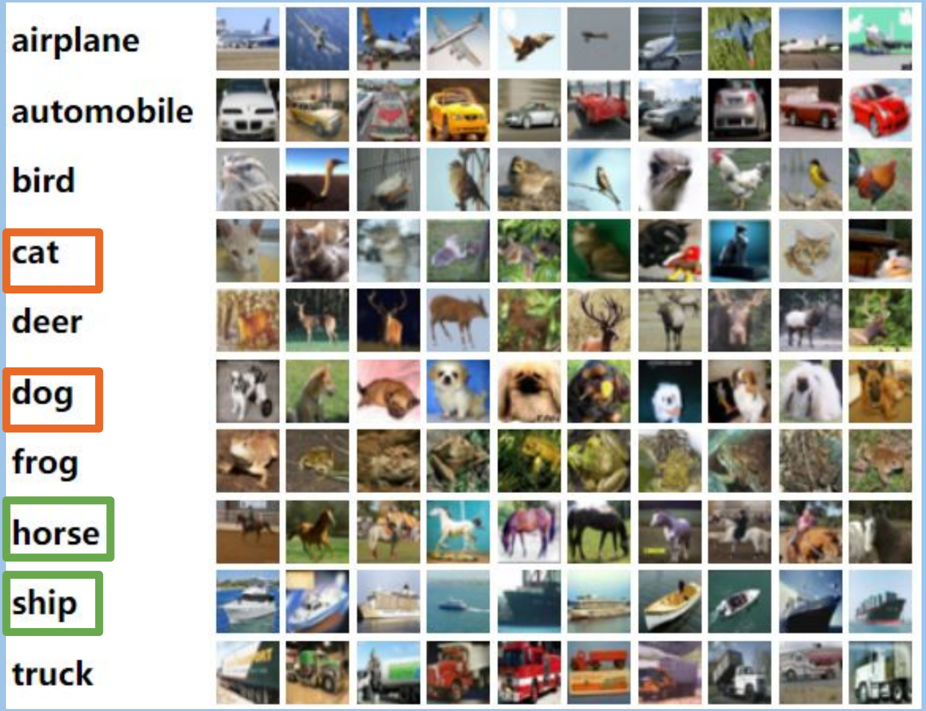
CIFAR10

Architecture and Dataset

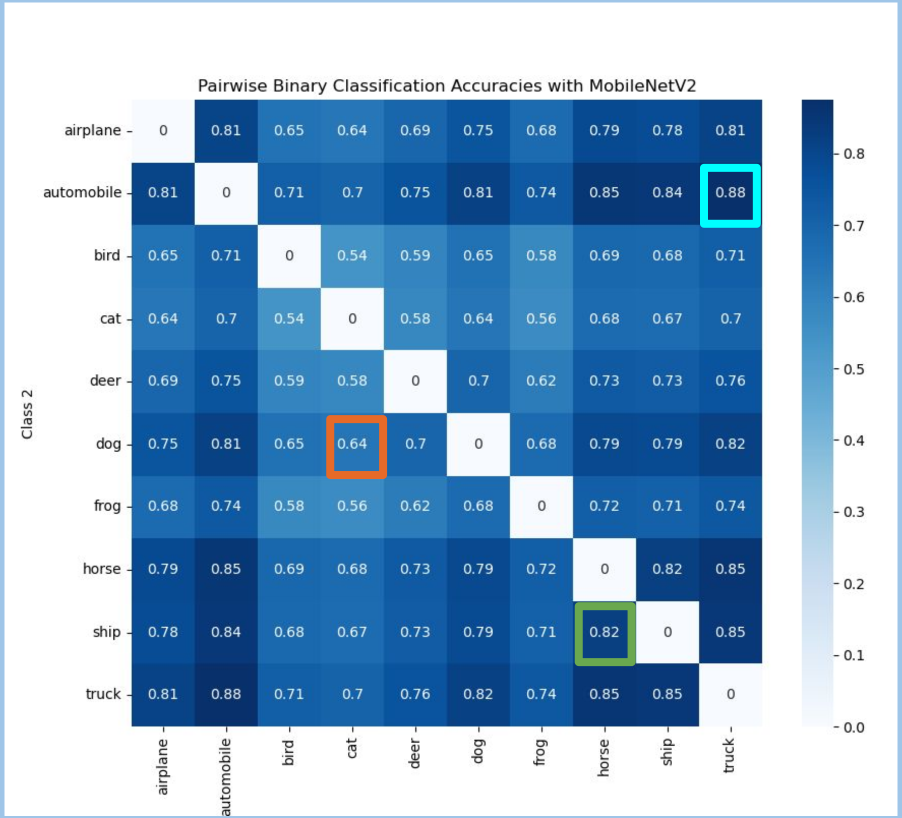


Choosing a linear classifier

Choosing binary subsets of the data



Choosing binary subsets of the data



Binary Subsets:

1. Dogs vs Cats (Hard)
2. Horse vs Ship (Easy)
3. Truck vs Automobile (??)

Final Approach

"Does a CNN's layer-wise class manifolds get more linearly separable the deeper you go?"

Algorithm:

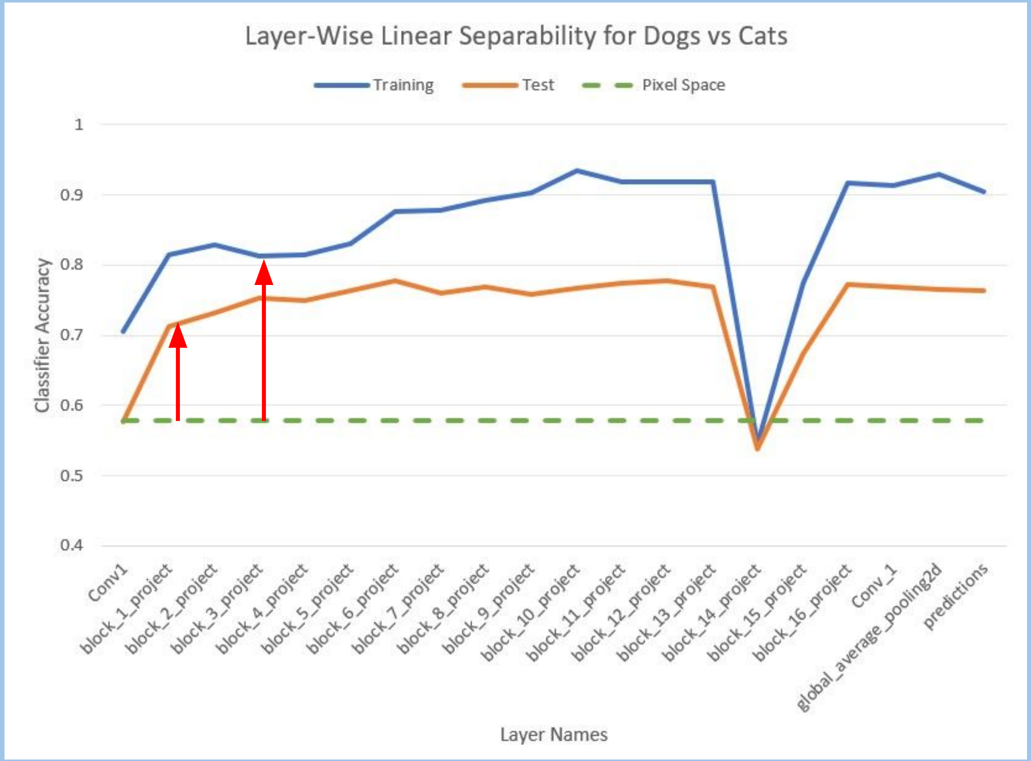
1. Obtain a CNN and dataset keeping in mind the Dimension problem
2. Train the CNN on the Image Classification dataset
3. For the 3 binary subsets of the dataset:
    a. Get layer-wise activations (manifold points) for two different objects
    b. Verify degree of linear separability for each layer activations

04

Results

Analysis and Anomalies

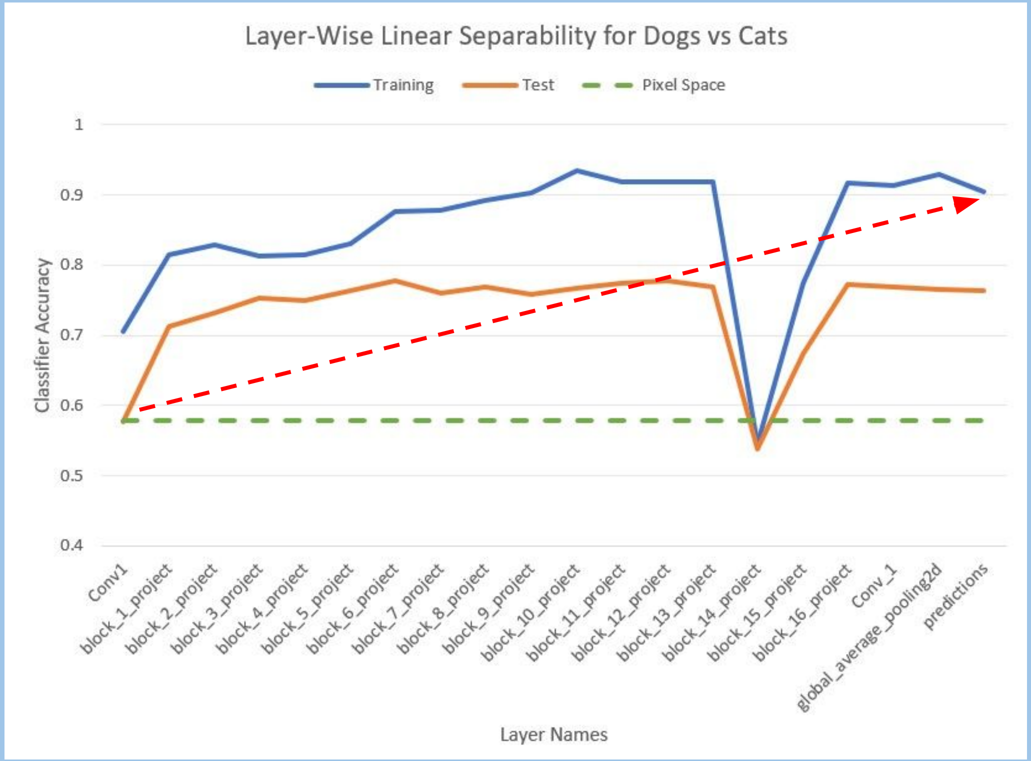
Initial Observations – Dog vs Cat



Observation 1:

Initial improvement in linear separability from pixel space

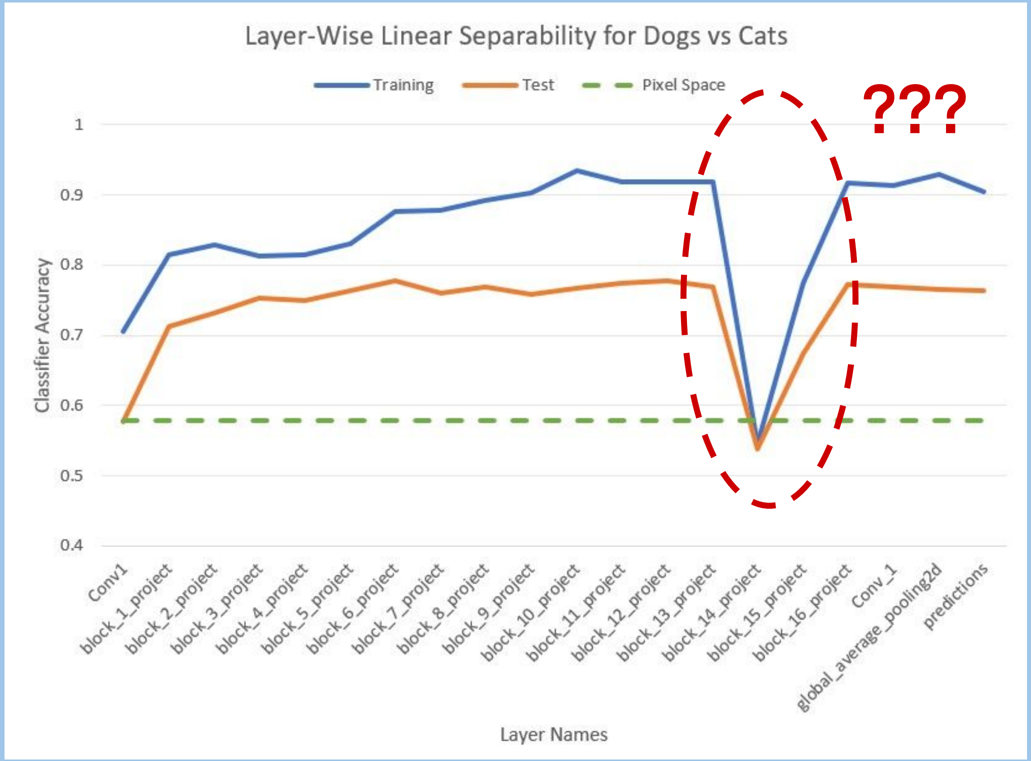
Initial Observations – Dog vs Cat



Observation 2:

Increasing trend in degree of linear separability

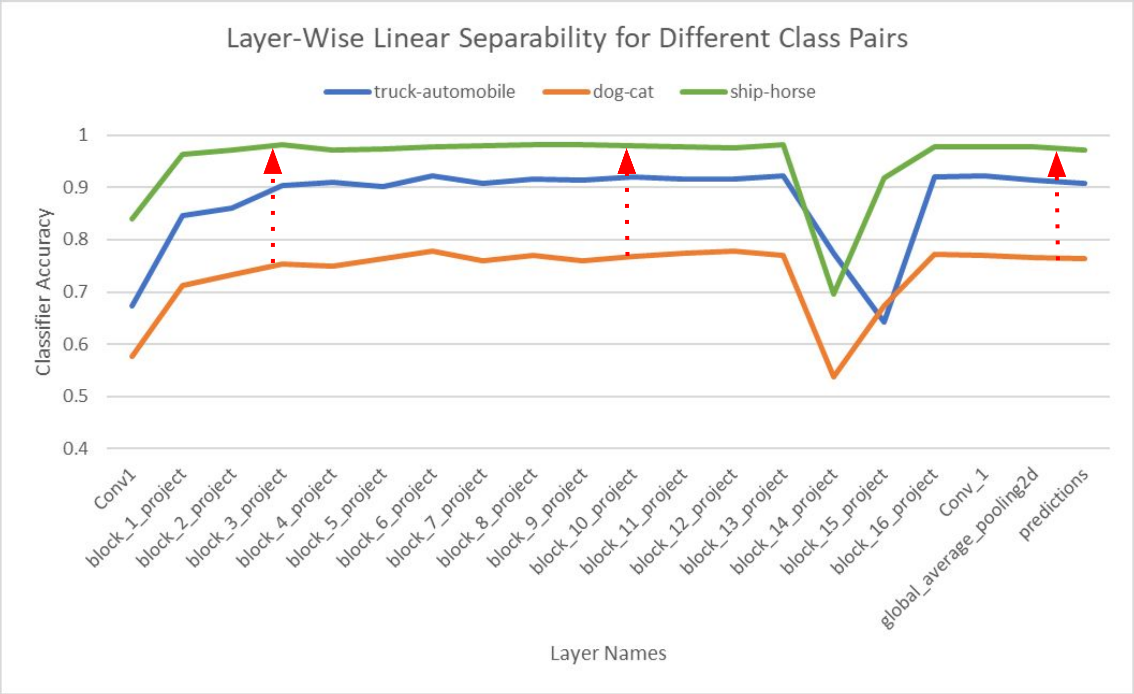
Initial Observations – Dog vs Cat



Observation 3:

Uncharacteristic drop in separability towards the top of the network

Observations for different binary data subsets

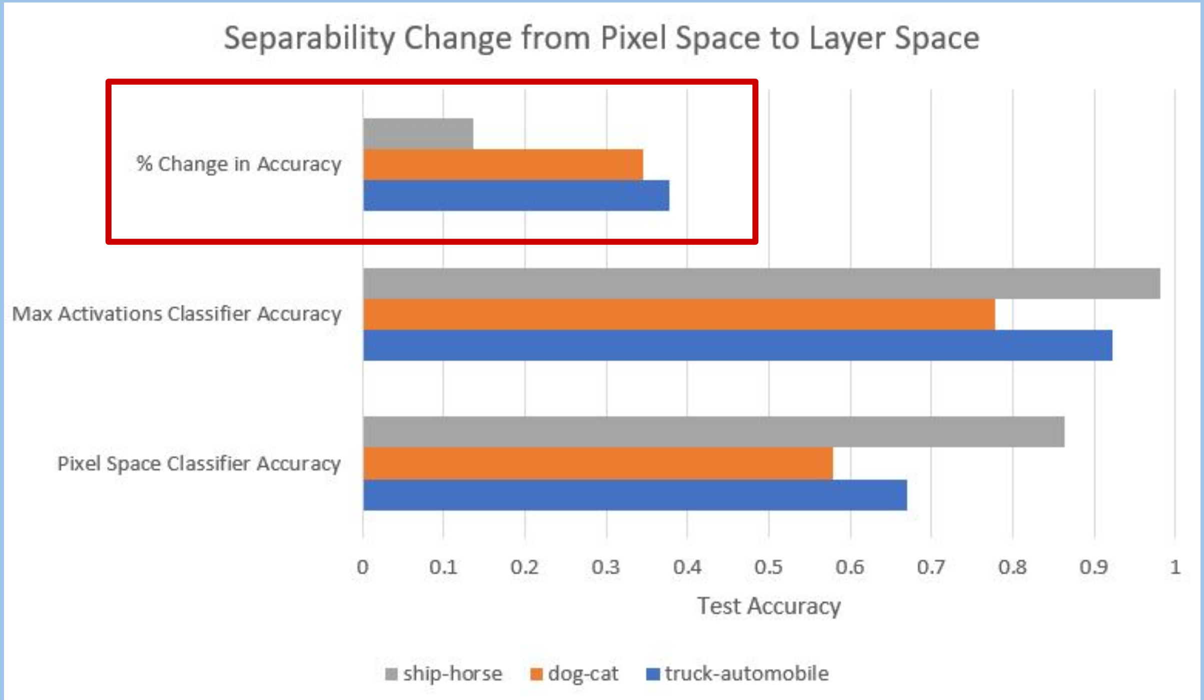


Observation 4:

"Easier" datasets achieve more separability in the layers

What about Truck v Auto?

Observations for different binary data subsets



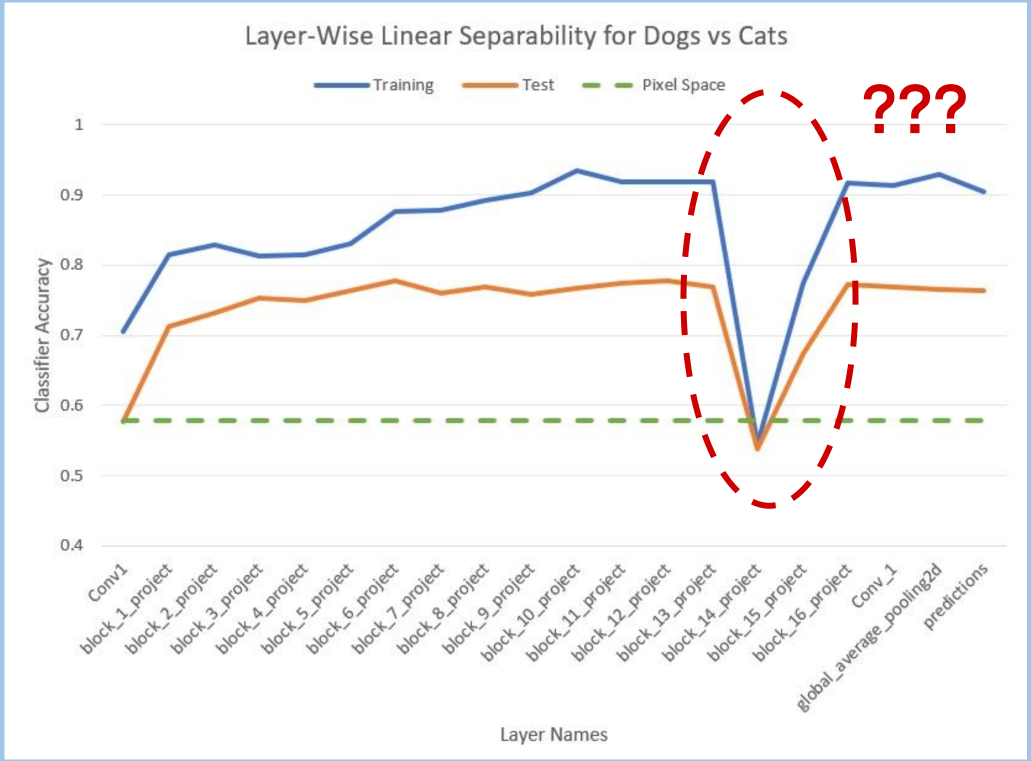
Observation 5:

Most learning has taken place for truck-auto

Observations for different binary data subsets

Learning decision boundaries is equivalent to learning to 'transform' representations into more linearly separable spaces!

The Uncharacteristic Drop



Possible reasons:

- Too many layers
- Change in Feature learning strategy

Needs to be further investigated

05

Conclusion

Takeaways and Future Work

Takeaways

- The first few layers of the CNN significantly improve linear separability
- The layers of the CNN improve linear separability with depth
- The ‘difficulty’ of dataset determines the degree of separability a CNN can achieve
- Learning decision boundaries == Learning to separate manifolds

Future Work to address Limitations

- Investigate the drop in separability further
- Verify results across more architectures and datasets
- Binary to 3-way, 4-way, n-way classification
- Class Manifolds vs Object Manifolds

Tying it all together – Thesis Directions

- Results suggest that CNNs process visual information in a similar manner — manifold separation
- Can we analyse behaviour of CNNs with the help of manifolds?
- Can we manipulate the manifolds of networks to design new behaviour?
- Simulating neural manifolds for tasks? Advance AI?

Thank You

Prof. Debayan Gupta, Ashoka University
Prof. Venkat Ramaswamy, BITS Pilani
Prof. Subhashis Banerjee, Ashoka University
Prof. Raghavendra Singh, Ashoka University