SMAI Project proposal

Team Name: Cassata

Team members:

- Abhijit Manatkar (2019101108)
- Kushal Kumar Jain (20191110001)
- Saravanan Senthil (2019101016)
- Tathagato Roy (2019111020)

Project Objectives:

Motivation

Since AlexNet won the ImageNet Challenge in 2013, Convolutional Neural Networks (CNNs) have become increasingly popular in various Computer Vision Tasks. Yet it is often unclear why CNNs perform as well as they do and how they learn to perform these tasks. Our lack of understanding of underlying mechanisms is an important bottleneck in designing better architectures and building better models.

In our project, basing on the work done by Mathew D Zeiler, Rob Fergus in their paper “Visualizing and Understanding Convolutional Networks”, we perform a series of experiments that hopefully would give us a better sense of what goes on under the hood of various CNN models.

We utilize what are called De-convolutional Networks to transform from the activation layers back to the image space which helps us visualize and understand what aspects of the image are the layers of the CNN focussing on.

We hope our understanding of the underlying mechanism of the various models helps us to develop a systematic way of designing different models for various tasks

Feature Visualization :

We aim to understand the feature maps learned in each layer by visualizing the patterns which cause the highest activations. These patterns would hopefully inform us what aspect of the image the feature map would strongly respond to.

Occlusion Sensitivity:

We aim to see if the CNN does in fact get inferences from the object in question and not from the context in the object's surroundings. This can be done by systematically occluding certain parts of the image and then looking at how the resulting image affects the end classification performance.

Feature invariance :

We explore the changes in maps that are caused by rotating, translating and shearing certain images. We plan to focus our attention on the first and the last layer of the model. The changes in the feature maps compared to the untransformed images would tell us how invariant the different layers are to these transformations.

Architecture Selection :

The ultimate aim of the project is to provide us a principled way of thinking about CNNs.A stronger understanding would allow us to select Architecture for a particular task in a more principled fashion than trial and error. In this project, we explore how to improve the general performance of the CNN's by making small changes to the Architecture, guided by our visualizations of the feature maps which tell us what features are learned in each layer.

Miscellaneous :

We will attempt to train our model in a suitable dataset(CIFAR or Imagenet) depending on the available time and resources and perform the designed experiments on them.

We will also be running the experiments on various pretrained CNN vision models and compare and contrast the differences between the various models(If there is sufficient time)

Correspondence Analysis :

Deep Models are opaque in the sense it is unclear how what exact features that they rely on to make their decisions. Do CNN's focus on the eyes or the nose while trying to detect faces? We are not explicitly inserting them in our design of the learning process.So are these objects parts being implicity learned? This is one of the questions that we aim to answer.

To achieve this we hope to design experiments that establishes correspondence between specific parts of the image and the resulting feature map vector.

One of the ways we do this is by seeing the changes in the features responses when certain objects or object parts within an image in occluded.

The difference in the feature responses gives us a concrete way of measuring change resulting from the masking operation. If the changes can somehow be understood to be systematic we can establish some correlation between specific features and the resulting outputs.

Miscellaneous :

We will attempt to train our model in a suitable dataset(CIFAR or Imagenet) depending on the available time and resources and perform the designed experiments on them.

We will also be running the experiments on various pre-trained CNN vision models and compare and contrast the differences between the various models(If there is sufficient time)

Project Deliverables:

- Implementation of all the models and boilerplate code on which we will run out experiments.
- A notebook with the specific experiments and results
- Output images that demonstrate the inner workings of a convolutional neural network
- A report describing our methods, the underlying theory, experiments, results, and conclusion.

Division of Work:

- Feature evolution during training - Tathagato Roy
- Boilerplate code - Saravanan Senthil
- Occlusion Sensitivity, Feature Invariance, Architecture selection - Abhijit Manatkar

- correspondence analysis, Model training - Kushal Kumar Jain
- Report writing - common

Timeline:

7 Nov - 17 Nov: Boiler Plate Code + Custom Model Training. We also hope to perform an initial set of experiments to see how everything is working.

18 Nov - 1st Week December: We hope to complete all our experiments during this time. We hope to have the final set of results and finish the final report detailing our methods, experiments, final visualizations and results, the conclusions, and intuition that can be drawn from our work.