Interactive Machine Learning and Visual Explanations

Track: Electrical Engineering & Computer Science - Data Science & Systems
Advisor: Prof. John Canny
Mentor: Biye Jiang
Team: Karishma Gulati, Jing Hong, Joseph Moon, Qingyang Zhao

Neural Networks - your complex yet best friends!

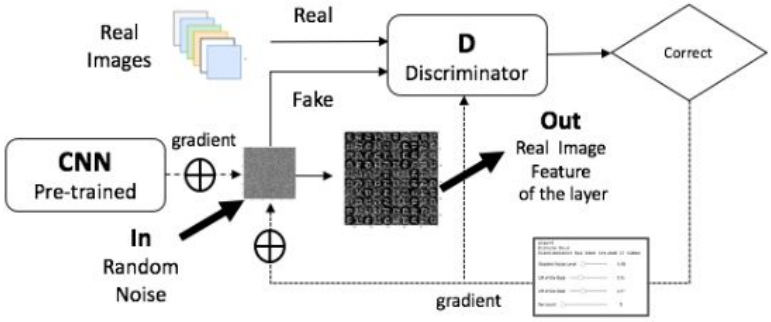
In recent years, neural networks have gained prominence as adaptable and robust problem solving algorithms for myriad tasks. However, as network complexity has grown, understanding the internal behaviour of networks has become a significant challenge. We aim to build a tool which will enable visualisation of the internal workings of neural networks while they are being trained. Users will be able to manipulate affecting-parameters in real time and see image representations of its behaviour.

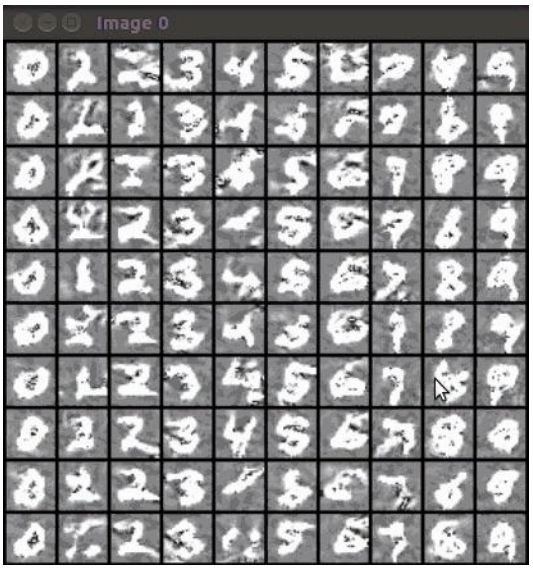
Let us help you understand them inside-out. But how?

First Step: (Activation Maximization)
Generate image representations for features.

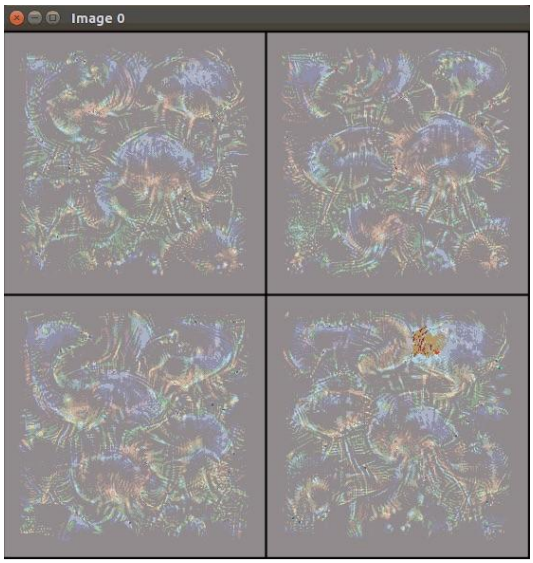
Second Step: (Discriminator)
Materialize these images as “human interpretable”.

Third Step: (Interactivity)
Interact with the discriminator weight and other parameters until the images make sense to you!





Visualization of neurons for handwritten digits (MNIST dataset)



Visualization of neurons for jellyfish (ImageNet dataset)

FUNG INSTITUTE for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING