CS 583: Final Project Proposal

Group Members

Name	Student ID	Email
Costa Huang	14314562	sh3397@drexel.edu
David Grethlein	12513707	djg329@drexel.edu

Citation

[1]: Simonyan, Karen; Vedaldi, Andrea; Zisserman, Andrew. *Deep Inside Convolutional Networks: Visualizing Image Classification Models and Saliency Maps*. ICLR, 2013.

[2]: Yu, Fisher and Koltun, Vladlen. *Multi-scale context aggregation by dilated convolutions*. ICLR, 2016.

Existing Implementation

We have found this repository with promising saliency map and neural activation visualizations already implemented. GitHub Repo

We hope essentially replicate their process on the MNIST dataset. kaggle

Abstract

There has been extensive research done in visualizing saliency maps to give human researchers (and tech enthusiasts) a better understanding what exactly a computer expects to appear in an image known to be linked to a class label [1]. Sophisticated convolutional neural networks are able to create filtered representations of image content at various resolutions in order to learn the 'high-level' and 'low-level' features that contribute to forming an image class label. While there has been a fair amount of attention paid to visualizing at the by-pixel basis given a certain class value, it is unclear if there has been any work comparing filters using different dilation factors and the impact of tuning on final class saliency maps. One possible outcome of using higher dilation factor would produce less blurry saliency maps since CNN with higher dilation factor preserves more spatial and structural information [2]. Our work aims to apply established procedures for generating class saliency maps at different kernel dilations.

Requirements

We plan leveraging existing implementations (may write from scratch too if time allows) of convolutional neural nets to process the MNIST dataset and visualize the different class saliency maps, as well as generating image-specific classification saliencies. By tuning different kernel dilation factors, we hope to draw analytical conclusions into the true effect of different receiver fields and pooling layers on forming our class understandings. ***Specifically*** the questions we hope to answer are:

- Is there a consistent way to *learn* the best dilation factor for a convolotional kernel?

- What is the effect of changing dilation factor when forming class representations (class saliencies)?

- Does dilation factor have a significant impact on classification accuracy?