

UNDERSTANDING INTERMEDIARY LAYERS OF A DEEP CONVOLUTIONAL NEURAL NETWORK USING NEURAL STYLE TRANSFER



Madhusudhan B A
WhirlData Inc. ©2018



1. Introduction

An understanding of image representation is a crucial component of any image understanding system. In particular interest here is the ability to understand image representations and significance in the various layers of a deep Convolutional Neural Network (CNN). CNNs have become the preferred to alternate to solve visual recognition / identifications problems that cannot be addressed through Computer Vision due to the wide variations in image characteristics that prevent a mathematical classification model.

An attempt is made here to answer questions of the nature as to what can we learn about image representation at different layers of a deep CNN.

How do different layers encode images, the nature of result of various of operations that can be applied across different layers and/or within filters of the same layer and the effect of altering hyper-parameters. We used an experimental approach to develop this understanding and for this purpose we used a network architecture inspired by Gatys et al1. developed for Neural Style Transfers(NST).

2. Charecteristics of the images

An image can be characterized by its "content" and "appearance". "Content" can be defined as those features that humans use to describe images (objects, shapes, actions) – "This is a picture of the Statue of Liberty", "That's the Great Wall of China in the background". These are also known as semantic features of an image. Until recently (a few years back), computers were not capable of using semantics for any kind of an image recognition, classification or retrieval tasks (at least such methods were not accessible to one and all and required expensive computing resources).

Traditionally low level image features derived from pixel values and relative differences between values of adjoining pixels were used for processing images and identifying patters that could be considered as those akin to what humans can recognize.

However, there was a large gap between the capabilities of computers and semantic features – also known as the semantic gap.

3. The role of Deep Learning & CNN

With advances in Deep Learning and Convolutional Neural Networks, it has become

possible for developing algorithms that are able to separate an image into its "content" and "appearance".

"Appearance", since we have not defined earlier and also known as the style of an image, is defined by different attributes such as texture, color schema, look and feel and anything else that describes the feel of an image. For instance, a painting created by using water colors will have several style definitions (by various people) and one such will be that of having the feel of water color painting. Some could say that the painting uses only primary colors and hence said to have a style of primary colors. Style is anything that maybe used by humans to describe an image. If we again use paintings as an example, a lithograph, a Tanjore art, a pencil art, an oil on canvas etc. will constitute a style.

Having the ability to split Content and Style allows for great innovation as it allows for the ability to recognize objects irrespective of any variations (as long as these variations preserve the objects identify in some manner).

One of the first uses of splitting content and style, thought not a commercial one, is to be able to transfer the style of one image to another. If we would like the contents of an image to be preserved, but represent the same image in another style, then the process of using CNNs/ Deep Neural Networks is called as Neural Style Transfer.

Neural Style Transfer requires a deep understanding of what goes on in several layers of a deep convolutional network usually containing several tens of millions or parameters (and in some cases 100 of millions of parameters).

Tapping different layers of the neural networks will yield different types of style transfers and experiments using NST serve as a basis for deeper investigation into what each layer of network does and can be used for. Due to the depth of the network and the plethora of parameters, experimental (and not just theoretical) work is required for this understanding and the result of such an experimental work opens up several new avenues for using CNNs effectively for various applications.

4. Neural Style Transfer

We developed a version of NST from a publicly available model (with due to credit to the Visual Geometry Group at the University of Oxford and Prof Andrew Ng who needs no introduction). We then customized the use of different layers effectively for various kinds of style transfers.

The model is based on the 19-layer VGG2 network sans the fully connected layers leaving behind a network with 16 convolutional layers and 5 pooling layers. We will not delve into the architecture here as the

reader can easily find references online to both the VGG 19-layer network as well the model inspired by Gatys et al.

We were interested in 2 parameters to start with – the number of iterations and the layer from which the style (style layer) representations are captured.

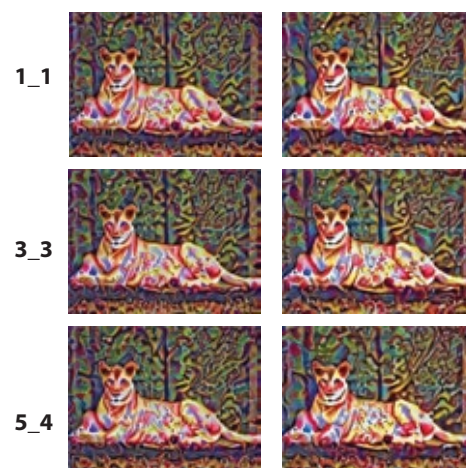
We studied the impact of using different style layers.

The resultant images (the picture of the lion "painted" in the style of the artwork) from different layers and 2 different values for iterations are shown above. Since the size of the images have been reduced to fit into this paper (and the differences may not be evident) – on careful scrutiny one can see that the quality of the images from layer 5_4 are far superior to the earlier layers - image structures captured by the style representation increase in size and complexity when including style features from higher layers of the network. This can be explained by the increasing receptive field sizes and feature complexity along the network's processing hierarchy.

In continuation of this work, over the next few weeks, we will also be experimenting on the relative weightage provided to content and style and their impact on image representations.



Layer 200 iterations 800 iterations



5. Style your own image

<https://whirlidatascience.com/whirlart/> showcases an experimental implementation of Neural style transfer that the reader can try for themselves.

6. Going forward

A slightly more complicated experiment would be on the use of various layers for content representation and the weightages provided to those layers and their impact on image representations.