Stephen Scarano
12 May 2021

Notes on Convolutional Neural Networks

I. Resources

A Comprehensive Guide to Convolutional Neural Networks: https://bit.ly/33BfiTh
Neural Style Transfer: Creating Art with Deep Learning: https://bit.ly/3w1SKal

II. Summary and Notes

Neural Style Transfer – Deep learning practice of composing images in the style one another

- Principle: define two distance functions, one describing the difference in content between the two images, *Lcontent*, and another describing the difference in style, *Lstyle*.
- **Input**:
  - A desired style image
  - A desired content image
  - An input image (initialized with the content image)

General Steps for Style Transfer

i. Visualize Data
ii. Basic Preprocessing and Data Preparation
iii. Create Loss Functions
iv. Create the Model
v. Optimize for the Loss Function

To best define and represent content and style representations within an image, we implement *intermediate layers* within the model:
Intermediate Layers: represent feature maps that become increasingly higher-ordered with deepness

Define **Content Loss**: a function that describes the distance of content from our input image, *x*, and our content image, *p*. Assume C_{nn} , a pre-trained deep convolutional neural network and some image, X, such that $C_{nn}(x)$ is the network fed by X.
Let $F^1_{ij}(x) \in C_{nn}(x)$ and $P^1_{ij}(x) \in C_{nn}(x)$ describe the respective intermediate feature representation of the network with inputs X and p at layer 1. The content loss (or distance) can then be described formally as

$$L^l_{content}(p, x) = \sum_{ij} (F^l_{ij}(x) - P^l_{ij}(p))^2$$

Perform *backpropagation* such that the content loss is minimized. Thus, the initial image is changed until it generated a similar response in a particular layer as the original image.

Define **Style Loss**: a function that describes the distance of style from our input image, *x*, and our style image, *p.* Follow the same principle as above but instead compare the Gram matrices of the outputs in place of the raw intermediate ones.

We may describe the style representation of an image as the correlation between different filter responses given by the Gram matrix G^l, where G^l_{ij} is the inner product between the vectorized feature map i and j in layer 1.
Perform gradient descent from the content image to transform it into one matching the style representation of the original image, generating a style for our base. In practice, minimize the mean squared distance between the feature correlation map of the style image and the input image. We may describe the contribution of each layer to the total style loss as

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{ij} (G_i^l j - A_{ij}^l)^2$$

Where G^l_{ij} and A^l_{ij} are the respective style representation in layer 1 of input image, X, and style image, A. N_l describes the number of feature maps, each of size M_l (equal to hight x width). Thus, total style loss across each layer is

$$L_{style}(a, x) = \sum_{l \in L} w_l E_l$$

This definition weighs the contribution of each layer by some factor w_l. In this case, all layers are equal such that $w_l = \frac{1}{\| L \|}$