DSP Final Project Team 20
Fourier Style Transfer

Members: Yao Hsiao(蕭瑤)

I. Problem Formulation / Scenario for Application

Neural Style Transfer (NST) performing image transformation has been an interesting topic since the proposal of [1]. The use of DNN in NST is novel in its capability to transform the style of one image into another one of sheer difference, but its computational cost is quite high to generate high quality styled images.

In this project, we want to use Fourier analysis and techniques in traditional signal processing to approximate the process of NST and apply the analysis to device filters to do image transformation in frequency domain. Specifically, to simplify the problem, we set *The Starry Night* by *Vincent van Gogh* as the style image, the intermediate objective is to find a regularity in the frequency spectrum, and the ultimate goal is to obtain filters that can transform any image to starry-styled ones, just like [Sup-Fig-1] done by DNN.

II. Problem Analysis

From the perspective of spectrum analysis, for an input image $img(x,y)$, the styled image $styled(x,y) = \sum f_i \otimes img(x,y)$ is assumed to be the result of convolving the input image with a set of filters, which represents the "style" of the style image, for example, the hallucinatory component of *The Starry Night* that distorts everything in the scene. To approach the problem, we assume that f_i are fixed. As mentioned earlier, the ultimate goal, if possible, is to obtain a fixed set of filters f_i, so first we need to establish the relationship between input image and styled image in frequency domain.

Taking Fourier transform of both side of equation of styled image and assuming that the application of filter is in regardless of input (for sake of simplicity), the problem becomes how to get a universal set of matrices Fs such that $FFT\{styled(x,y)\} = S = \sum F_i \times I = (\sum F_i) \times I = F \times I$, where the multiplication is element-wise and I is the spectrum of input image. [Sup-Fig-2]. This means the styled images are essentially multiplication of two matrices. To analyze the potential relationship between I and S so as to obtain F, we first collect a group of pictures from MS-COCO dataset and the styled version using the DNN set up on Google Colab [2]. Our dataset is of size 24. [3]

An essential hypothesis in previous derivation is that the usage of filters F_i is universal for all images, which may not hold true since there is quite obvious difference in how *van Gogh* treated different objects and scenes. Therefore, the problem, in a more complete sense, should be $S = \sum F_k \times I \times a_k(I) + F_{common} \times I$, where $a_k(I)$ are filters dependent on input images.

However, there must still exists a common filter (equivalently the common style) that are applied to all images, that is the naïve hypothesis still holds partially, and to approach the problem, we, starting in the next section, will first focus on the universal F first.

III. Methods

I. Inspection on the validity of common filter F_{common} by statistics
Instead of using filter bank to decompose the signal into sub-band for analysis (an initial trial in Execute/src/ batch_ftbank.m) as mentioned in proposal, it turns out that sub-bands signal of images included too much information for analysis, and not a convenient tool for aggregate analysis on a bunch of data, and thus we take a more statistic approach.

a. Plotting to inspect potential regularity in the spectrums of styled images and input ones, such as one on one comparison of statistics (such as mean, histogram) on amplitude and phase of S and I themselves, $S./I$, $S-I$. [Sup-Fig-3]. In a more detail sense, we resize all images to 64 by 64 and only inspect one channel of the images.

b. Since $\mathrm{S} = \mathrm{F} \times \mathrm{I}$ the multiplication of matrices is element-wise, if such F exists, then each frequency component should be in a linear relationship, that is $S(x,y) = \mathrm{F}(\mathrm{x},\mathrm{y})I(x,y)$ $\forall\ x,y$. Therefore, by statistics, we use correlation coefficient to inspect such trend.

c. A naïve manual filter from the regularity observed above and see how much of the style can be achieved with this filter comparing to the result of DNN on the same image.

II. Toward a more detailed style filter to approximate $\mathrm{S} = \sum F_k \times I \times a_k(I) + F_{common} \times I$

a. The problem is twofold, the design of filters F_k, and design of $a_k(I)$. Because of limited amount of time, we will simply let F_k taken from the dataset. That is $F_k = S_k./I_k$, where (S_k, I_k) are from dataset.

b. Now we can possibly focus on $a_k(I)$. By $\mathrm{S} = \sum F_k \times I \times a_k(I) + F_{common} \times I = F'(I) \times I$, we first inspect the relationship between I and $S./I$

c. Devise single measure of image I so to quantize the relationship.

d. Use the aggregate of F_k and a_k on the new image

III. Contingency plan

If, the previous two approaches do not yield good enough result, we will turn to a modified version of hybrid image proposed by [8]. The hypothesis will then be using low pass filter G, and high pass one 1-G, $\mathrm{S} = \sum F_k \times I \times a_k(I) + F_{common} \times I = G \times I_1 + (1-G) \times F$, where I_1 is the image to be styled and F may contains the style. This workaround may work only for *The Starry Night,* for that this image has quite a lot of high frequency components, and thus using high pass filter to convey the style into the image may thus possibly work while retaining the low frequency component of I_1.

IV. Result Analysis

I. Inspection on the validity of common filter F_{common} by statistics

a. Initial trial on filter bank in decomposing the signal (in Execute/src/filter2d_gen.m) and its application on the images to see (in Execute/src/batch_ftbank.m). Again, it is not quite a convenient approach.

b. As mentioned in method I. a., we tried out different comparison of statistics on $S./I$(element-wise) and $S-I$, to validate the common mode linear relationship (if there is common mode, then the plotting should approximate horizontal line). The amplitude and phase result (in Sup-fig-4 and 5 respectively), where each data point in a plot is the same frequency component of the whole dataset , x is the amplitude of that frequency component of the raw image , while y is the amplitude or phase of $S./I$. All numbers are normalized against the dataset. We only randomly select some of the frequency component to make the plot and tried linear regression on the plot. The figures on the title of the plot are (R-square, standard deviation of y quantity of that frequency component, and number of points).

From the result, for amplitude part, the R-square are all quite low suggesting no clear linear relationship, but the quantity of standard deviation of amplitude for most frequency component are within a small range of 0 and 1. This means change on amplitude of each frequency component may have a common mode for most of the images. To validate this, we simply obtained an amplitude filter from any $S./I$ and applied it to see the effect, and it turns out the images are all quite styled with the distorting effect [Sup-fig-6, and Execute/src/amptransfer.m], but lack the detailed distorting pattern [Sup-fig-7]of *the starry night.*

For phase, there seems no obvious common mode filter that can be applied to all images, for that the R-square are all quite low, and the data distribution are all quite random with large standard deviation. It turns out that a common mode filter can only possibly and partially work for the amplitude part but not for phase. However, from above methods and results of spectral plotting [Sup-fig-3], phases are much more important to create the targeted style.

II. Toward a more detailed style filter to approximate $S = \sum F_k \times I \times a_k(I) + F_{common} \times I$

a. As mentioned in method II. a.&b. , we focus on the relationship between I and $S./I$. Initially we first try to device single measure of image I so to quantize the relationship. A common thing to do in statistics is the use of histogram. Therefore, in the left-most plot of spectral analysis (Sup-fig-3), we plotted the histogram of amplitude of I, and intuitively used the shape histogram as a measure of similarity of images. From a quick observation, we found that the more similar the histogram between two Is ($I_1$and I_2) , the application of ($S_1./I_1$) on I_2 yields better effects. Such as in [Sup-fig-8], the histogram of image A is more similar to image B than to image C, and applying the filter ($S_B./I_B$) on I_A does yield a better quality comparing to result of applying C, in which all things are blurred.

a. As observed in previous point, we think there may be a relationship between the histogram and filters $S_k./I_k$, so we plot out the mean of amplitude of them [Sup-fig-9]. It turns out that there may be a linear relationship, but a closer inspection on the plot shows that a majority of data points lie in a horizontal line and some of the point may be outliers. In conclusion, the single measure of histogram is not so sufficient in the task and requires better measure to represent these spectral.

   In conclusion, a single measure is hard to find, that is the search of $a_k(I)$ is not so successful, but we do confirm that filters dependent on the images are definitely needed to achieve our ultimate goal.

b. From the previous result and [4], we then believe before applying a change to a frequency component, we should inspect the existence of pattern in image just like in the DNN approach the circle in the image may be styled with a circle-like distortion. But to build such analysis of spectral pattern may be too complex, and thus we have yet come up a good approach in using this observation.

III. Contingency plan [8] hybrid image
Suggest by [8], a hybrid image $I = G \times I_1 + (1 - G) \times I_2$. In our case, we will let I_1 be the image to be styled, and I_2 will be either a filter or an image.

a. I_2 be a filter extrapolated from data (S_k/I_k) : the results are in [Sup-fig-9], and implementation is in Execute/src/test_siggraph.m, which can be tested with other k in dataset. While we do see some distortion toward the style, the effect is not so obvious.

b. I_2 be the starry night, just like [8]: the results are in [Sup-fig-10], and implementation is in Execute/src/test_siggraph2.m. The effect of starry night is obvious, but the drawback is you are seeing the original image and starry night at the same time, instead of a real styled image.

V. Conclusion

In this project, we use traditional signal processing way to approach the task of style transfer. By using the paired images generated by DNN approach, we assumed each of them is filtered by a common filter. We first carry out Fourier analysis on the images and try to find the regularity of the filters. A statistic on these filters' amplitude yields that amplitude change shows a more universal tendency while the phase, which is more important in the task, yields no simple and obvious trend. To sum up, we could not find a single style filter independent of images. Therefore, we then turn to a more realistic hypothesis, and through a couple of qualitative analysis and comparison, found that the input image spectrum does effect change applied to different frequency components, but we have not effectively derived out the relationship.

VI. Contribution

A. Contribution of Method

The application of statistics in digital signal processing are mostly what we devised, and only hypothesis derives from basic ideas of 2D FFT and traditional digital processing on images [5].

B. Contribution of Implementation

- The implementation of DNN is referenced from [6]
- 2D filter bank [Execute/src/filter2d_gen.m] is referenced from MathWork doc [7]
- Execute/src/siggraph.m idea from [8]
- all others implemented by us

VII. Work Division

N/A. Team 20 is single-person team.

VIII. Reference

[1] Johnson, Justin; Alahi, Alexandre; Li, Fei-Fei (2016). "Perceptual Losses for Real-Time Style Transfer and Super-Resolution"

[2] https://colab.research.google.com/drive/1VC0gcmM1pmq50qCRBgkJvULefxDkrC6b

[3] https://github.com/yaohsiaopid/dspfinal/tree/master/img images ending with '_rs.png' are input, and those ending with '_starry.png' are styled

[4] https://people.eecs.berkeley.edu/~efros/courses/LBMV07/presentations/0208Gist.pdf

[5] DG Manolakis, VK Ingle. "Applied digital signal processing: theory and practice"

[6] https://github.com/titu1994/Neural-Style-Transfer.git

[7] https://www.mathworks.com/help/images/ref/fwind1.html

[8] A. Oliva, A. Torralba, P.G. Schyns (2006). Hybrid Images. ACM Transactions on Graphics, ACM Siggraph, 25-3, 527-530.