Utilizing Fourier Transformations for the Detection of Fake Images

Yunho Kim
Gwangju Institute of Science and Technology
Gwangju, Korea
youknowyunho@gm.gist.ac.kr

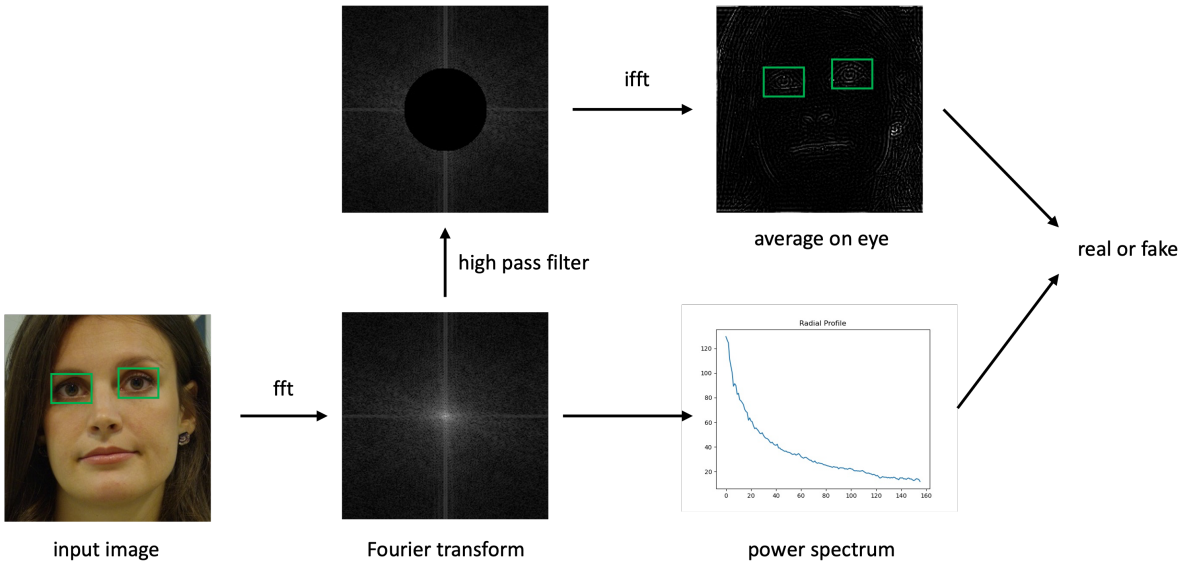

Figure 1. Overview of the proposed method.

Abstract

In the era of digital media, various face manipulation technology, such as face swap, age manipulation, or talking head, has emerged. This report [1] aims to identifying fake images using Fourier Transformations, a mathematical method widely used for signal analysis. While previous studies have employed pixel-based or feature-based methods for fake image detection, this research applies Fourier Transformations to employ the information from freqeuncy domain.

Our methodology involves decomposing images into their frequency components through the application of the Discrete Fourier Transformation (DFT). By examining the frequency spectra, we can detect difference in frequency domain, that is produced wile image being synthisized. This approach proves to be particularly effective in detecting fake images that appear visually flawless to the human eye.

This research not only contributes to the growing field of digital image forensics but also has significant implications for combating misinformation and maintaining the integrity of visual media in our digital age. Future work will aim at refining this approach for real-time applications and extending its applicability to video forgery detection.

1. Introduction

With the advancement of computer vision, technology to synthesize fake facial images has been improved rapidly. Such developments have profound implications for various domains, including journalism, law enforcement, social media, and personal privacy, thereby making the detection of

[1] code available at https://github.com/YunhoKim21/FT-FakeImageDetection

synthetic images a matter of significant moral and social concern.

This work presents an approach that leverages the power of Fourier Transform – a mathematical technique extensively used in image processing – to detect synthetic images. Fourier Transform allows us to analyze an image in the frequency domain, providing unique insights into its structure, and consequently, the ability to discern between genuine and manipulated content.

2. Related Works

2.1. Face Manipulation

Face manipulation, a research field that investigates to synthesize of facial images with desired properties, has gained a lot of attention since the founding of computer vision. Talking head generation, reaging, and face swap are subdivisions of this research field. As they are all generative models, their development have heavily relied on the development of generative models. For example, when Generative Adversarial Networks (GAN) [2] was introduced, many types of research employed GAN architecture for their study [1] [7]. When the next generation generative model, Denoising Diffusion Probabilistic Model(DDPM) [4] emerged, many follow up researches applied DDPM in face manipulation [5].

As the backbone of fake facial image generation keep changes, it is important to gain the ability to detect fake images, produced by state of the art methods.

2.2. Discrete Fourier Transform

Discrete Fourier transforms in 2D are defined as follows:

$$F(u,v) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)\text{exp}[-2\pi i(\frac{mu}{M} + \frac{nv}{N})] \quad (1)$$

Implementing the discrete Fourier transform directly is technically feasible. However, such an approach would necessitate using four nested iterative loops, leading to computational inefficiencies. This research will use another representation of the Fourier transform, which we will derive here.

For simplicity, let us define two functions, which will later be regarded as matrices as two integers are inputs.

$$A(n,v) = \text{exp}(-2\pi i\frac{nv}{N}) \quad (2)$$

$$B(u,m) = \text{exp}(-2\pi i\frac{mu}{M}) \quad (3)$$

Then the discrete Fourier transform (Eq.1) can be rewritten as,

$$F(u,v) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)A(n,v)B(u,m) \quad (4)$$

However, using the definition of matrix multiplication, we can further simplify as,

$$F(u,v) = \sum_{m=0}^{M-1} (f \cdot A)(m,v)B(u,m) \quad (5)$$

Using the definition of matrix multiplication once again, we finally get

$$F(u,v) = (B \cdot f \cdot A)(u,v) \quad (6)$$

Hence we can conclude that, Fourier transform can be represented as,

$$F = B \cdot f \cdot A \quad (7)$$

, where A and B are defined at Eq.2 and Eq.3. The inverse Fourier transform can be simplified in a similar manner.

3. Method

3.1. Implementation of DFT

Discrete Fourier transform(DFT) and Inverse discrete Fourier transform have been implemented using Eq.7. NumPy [3] module was used for fast computation. See function DFT2D and IDFT2D for implementation details. We use DFT and IDFT to analyze the image in the frequency domain, which will be explained in the following sections.

3.2. Azimuthal Averaging

Azimuthal averaging, also known as radian averaging, averages, the value of some image in a concentric circle. Since the center of the Fourier transformed image represents low-frequency components, Azimuthal Averaging is often used with Fourier transform to investigate how many high-frequency components the image has.

In means of fake image detection, it is well known that some GAN-based generative model struggles to synthesize high-frequency details. Hence we will average the Fourier-transformed image and compute the magnitude of high-frequency components. We hypothesize that the lack of high-frequency components indicates that the image is fake.

To obtain numerical criteria and prove our hypothesis, we use FaceForensics++(FF++) [6]. The dataset consists of real and fake facial images, and is widely used for fake image detection.

Specifically, we random sample 500 images from real and fake image sets, and then obtain the spectrums for each

images. Finally, we compare spectrums of our dataset with those from FaceForensics++ to decide if the image is real or fake.

3.3. High Pass Filtering

High pass filtering has been implemented by eliminating all frequency components within a circle. Fig.2 visualizes high pass filtering.

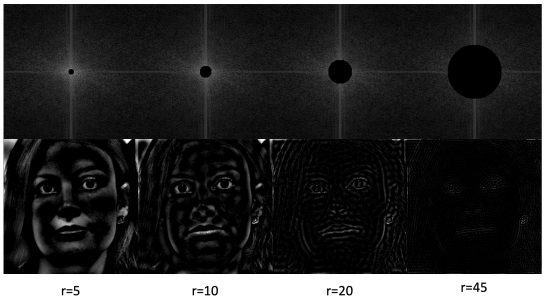

Figure 2. Visualization of high passed filtering, by different radius. Upper row represents high pass filtered image in frequency domain, and the lower row corresponds to the inverse Fourier transform of the upper row.

Although many different choices for filtering radius are possible, in this work, we choose r = 45(pixels) heuristically. We use high-pass filtered images in the next section to analyze how many high-frequency components are in an image.

3.4. Classification by Eye Activation

The eye is the region in the face that contains great high-frequency details due to its complex components, such as eyebrows, eyelashes, and iris. At the same time, GAN-based image generation methods [7] struggle to synthesize high-frequency details due to the feed-forward architecture. Hence we can hypothesize that examining the high-frequency components around the eye could be a decent criterion.

Based on this idea, we define "eye activation", which is the average intensity of high passed image, near the eye region. Eye activation is formulated at Eq.8. $I_{highpass}$ denotes the high passed image, and M_{eye} denotes the eye mask, where the intensity is 1 near the eye and 0 otherwise. We obtain eye masks using a pretrained face detection network [8]. We hypothesize for now that the low eye activation indicates that the image is fake. See Sec.4.2 for experiment results.

$$\text{eye-activation} := \frac{I_{highpass} \odot M_{eye}}{|M_{eye}|} \tag{8}$$

4. Experiments

4.1. Classification by Spectrum

4.1.1 Spectrum of FaceForensics++

Spectrums of 1000 images from the FaceForensics++ dataset are depicted on Fig.3. The number of fake images and real images is both 500. Each spectrum is drawn as a faint line. Also, the two think lines represent the mean of the spectrum. Blue and red lines represent real and fake images, respectively.

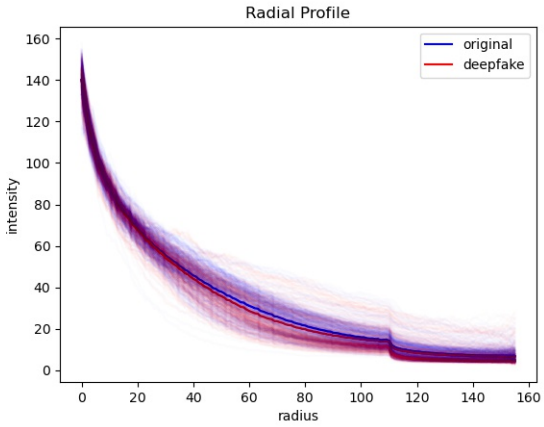

Figure 3. Spectrums of 1000 real and fake images from FF++ dataset.

We can notice that fake images have lower intensity in the power spectrum. This proves that our hypothesis in Sec.3.2 is right. However, the gap between two classes is small, but the variance of the spectrum is relatively high. Hence inconclusive decision may occur when deciding with power spectrum.

4.1.2 Spectrum comparison with our data

Spectrums from our dataset and the average spectrum from the FaceForesnics++ dataset is plotted below (Fig.5). Note that all images were resized to 224x224 size to make an equal comparison.

Spectrums of image number 1, and 2 are way below the average, which makes it reasonable to predict that they are fake. Also, for a similar reason, image number 3 can be classified as real. However, the spectrums of other images are located near the mean, which results in inconclusive results. Table 1 summarizes the prediction results.

4.2. Classification by Eye-Activation

In this section, we define eye activation as the average intensity of a highpassed image, only at the eye region. Eye activation represents how much high frequency details are

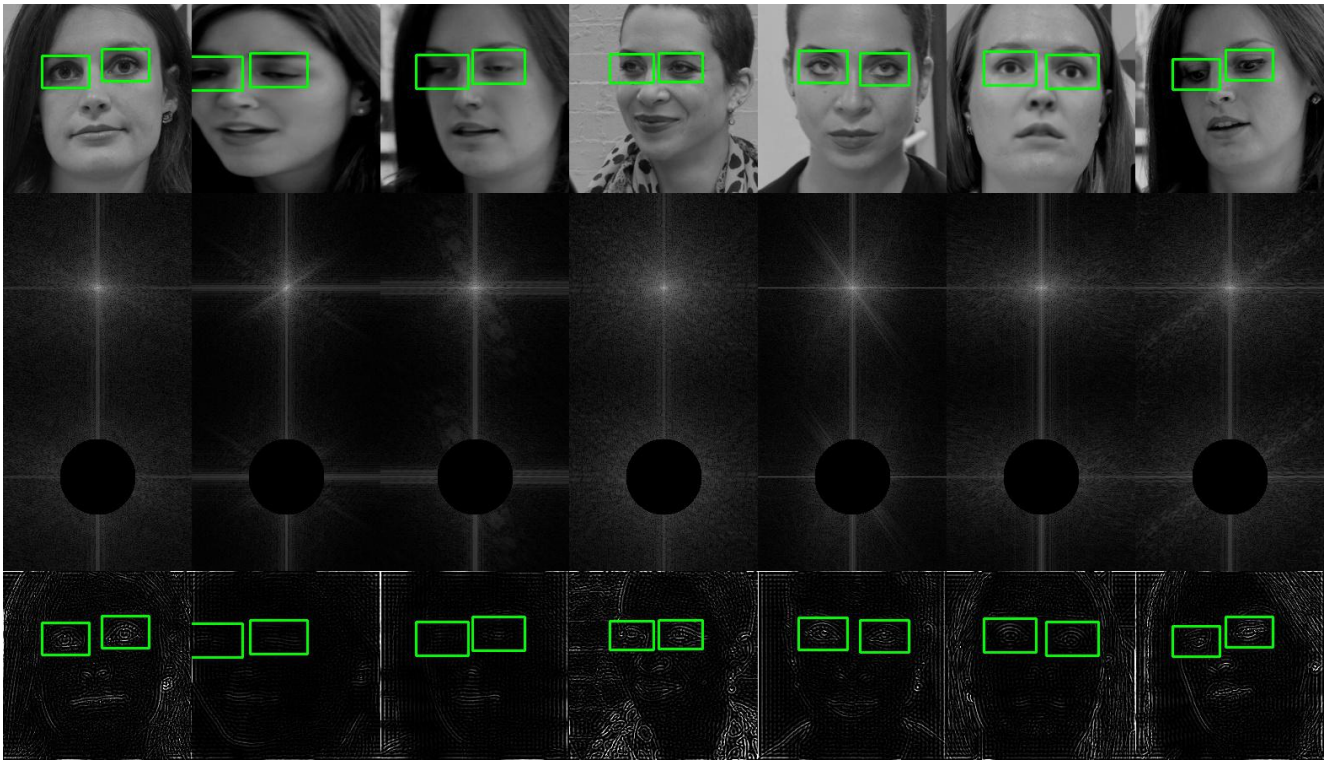

Figure 4. Visualization of images from our dataset. The first row shows resized, grayscaled images with eye detection. Resize is required for equal comparison between images. The green boxes represent the eye caught by the eye detection network [8]. The second and third row represents Fourier transformed images and high-passed versions. The last row visualizes the inverse Fourier-transformed image of the third row. The intensity of last row is manually scaled 10 times for better visualization.

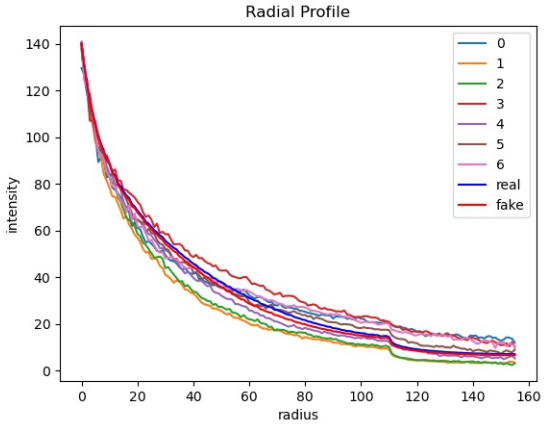

Figure 5. Spectrums from our dataset compared to the ones from FaceForensics++.

contained in the eye region. We hypothesized before that synthesized image will lack high frequency details near eye. This hypothesis will be proved at Sec.4.2.1. Also the hypothesis will used to classify our dataset, at Sec.4.2.2.

Index	Prediction
0	I
1	F
2	F
3	R
4	I
5	I
6	I

Table 1. Prediction Results for Image Data, by spectrum analysis. F, R, I represents fake, real, and inconclusive.

4.2.1 Eye-Activation from FaceForensics++

Mean and standard deviation for 1000 images in FaceForensics++ is summarized at Tab.2.

Image Type	**Mean**	**STD**
Fake Images	2.599	1.168
Real Images	4.461	2.305

Table 2. Mean and Standard Deviation of eye activation

As anticipated, fake images scored less eye activation

compared to real images, quantitavely proving that our hypothesis was right. Also, the gap between fake and real images is significant enough to classify most of images.

4.2.2 Eye-Activation from our data

Tab.3 summarizes eye activation from our dataset. It is noticeable that there is a significant gap between real and fake images.

Index	Eye Activation	Prediction
0	5.723	R
1	0.723	F
2	1.525	F
3	5.873	R
4	4.242	R
5	2.795	F
6	4.564	R

Table 3. Prediction Results. F, R, I represent fake, real, and inconclusive.

Visualization of eye activation is depicted in Fig.4. Looking at the bottommost row (high pass filtered image), we can immediately notice that image of index 1, 2, and 5 seriously lacks high-frequency details, especially in the eye region.

4.3. Application against State of the art Face Generation Models

As eye activation plays a great role in discriminating fake images, a question arises if it can discriminate synthesized results from state-of-the-art generation models. While GAN based models struggle to synthesize high-frequency details, Denoising Diffusion Probabilistic Models(DDPM) [4] emerged as a competitor of GANs [2], showing impressive performance. Hence we apply our eye activation test on modern DDPM based face generating works [5].

The results of eye activation test against diffusion models is summarized at Tab.4.

Image Type	**Mean**	**STD**
Diffusion based fake image	6.104	0.491
Real Images	4.461	2.305

Table 4. Mean and Standard Deviation of eye activation, against modern face generation networks [5].

Surprisingly, diffusion-based face generation network scored greater eye activation than the real images. Although we could leverage the gap between those, it could easily be circumvented by simple methods(e.g., blurring on eye region). Hence we conclude here that our method is not suitable to discriminate results from state of the art generation methods.

5. Conclusion

The final prediction is summarized at Tab.5. Despite that some results were not classified by spectrum analysis, eye activation test provided us full predictions with high confidence. Also, there was no case such that power spectrum analysis and eye activation gave different predictions, which makes our prediction even more credible.

Index	Spectrum	2D IFT	Final Predictions
0	I	R	R
1	F	F	F
2	F	F	F
3	R	R	R
4	I	R	R
5	I	F	F
6	I	R	R

Table 5. Final prediction Results. F, R, I represent fake, real, and inconclusive.

In conclusion, the application of Fourier Transform in image analysis has proven to be an effective tool in detecting manipulated and fake images. This study has showed that by decomposing an image into its frequency components, it is possible to reveal traces of image synthesis that are not immediately visible to the human eye. In particular, noise pattern inconsistencies, often introduced during image manipulation, can be effectively highlighted using the Fourier domain.

However, we also found that our model fails to discriminate for state-of-the-art diffusion-based face generation networks. Hopefully future works applying Fourier transform might detect these.

References

[1] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. Simswap: An efficient framework for high fidelity face swapping. In *MM '20: The 28th ACM International Conference on Multimedia*, 2020. 2

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 2, 5

[3] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser,

Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. 2

[4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. 2, 5

[5] S. Cho J. Seo J. Nam K. Lee S. Kim K. Lee K. Kim, Y. Kim. Diffface: Diffusion-based face swapping with facial guidance. *Arxiv*, 2022. 2, 5

[6] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. FaceForensics++: Learning to detect manipulated facial images. In *International Conference on Computer Vision (ICCV)*, 2019. 2

[7] Yuhan Wang, Xu Chen, Junwei Zhu, Wenqing Chu, Ying Tai, Chengjie Wang, Jilin Li, Yongjian Wu, Feiyue Huang, and Rongrong Ji. Hififace: 3d shape and semantic prior guided high fidelity face swapping, 2021. 2, 3

[8] Jia Xiang and Gengming Zhu. Joint face detection and facial expression recognition with mtcnn. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, pages 424–427, 2017. 3, 4