Ukrainian Catholic University

Faculty of Applied Sciences

Data Science Master Programme

Frequency transformation for image recognition and inpainting

Linear Algebra final report

Authors:
Igor Babin, Kostiantyn Verhun, Iryna Pastukhova

5 March 2023

Contents

1 Abstract

In this project, we explore the use of frequency transformations, specifically Fourier and wavelet transform, to improve image recognition and inpainting in Convolutional Neural Networks (CNNs). The motivation behind this work is the fact that images can be decomposed into their frequency components, which can then be used to extract important features that are readily apparent in the spatial domain. We first give a brief review of the problem and related works, then discuss the mathematical foundations of Fourier and wavelet transforms and the properties that make them suitable for image processing, and then experimentally show that adding frequency transformation can improve the accuracy of the model. The final code is available here https://github.com/igor185/frequency-for-receptive-field

2 Introduction

Convolutional Neural Networks (CNNs)[1] have been widely used in image processing tasks such as classification[2] (assigning some class to an image e.g. dog vs cat) and inpainting[3] (restoring masked region) due to their ability to extract relevant features from images. The receptive field[4] of a CNN plays a crucial role in its ability to recognize and distinguish objects in images. A larger receptive field allows the CNN to capture more context and spatial information from the image, which is important for accurate classification and inpainting. In this project, we explore the use of frequency transformations, specifically Fourier and wavelet transform, to increase the effective receptive field of the CNN model and improve its performance in these tasks.

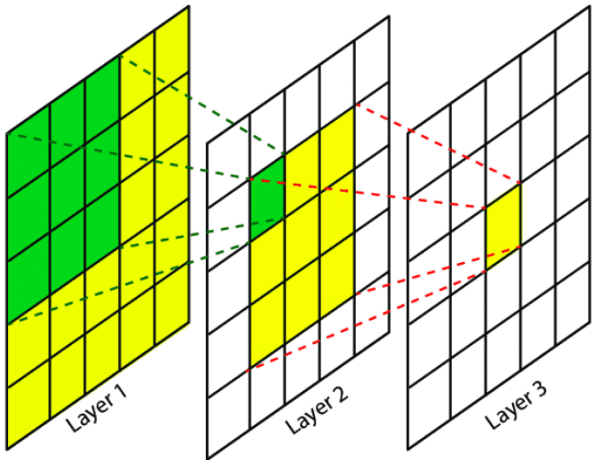

Figure 1: Receptive field for 3 layers of CNN

Images can be represented in both spatial and frequency domains. The spatial domain represents the image as a two-dimensional grid of pixels, where each pixel corresponds to a particular intensity value. The frequency domain, on the other hand, represents the image as a sum of sinusoidal waves of different frequencies and amplitudes. The Fourier and wavelet transforms are used to decompose the image into its frequency components, which contain information about global and local features of the image.

In image processing tasks, such as classification and inpainting, it is important to consider the entire image as a whole rather than just individual pixels. Each frequency component in the frequency domain contains information about the entire image, as opposed

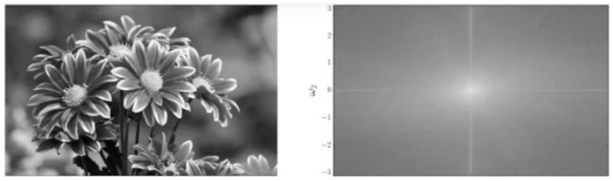

Figure 2: Example of image in special domain and in frequency domain

to just local features in the spatial domain. This is because the frequency components capture the patterns and structures in the image that repeat over larger regions, whereas local features are typically confined to smaller regions. By incorporating frequency transformations into the CNN model, we can capture this global information and improve its ability to recognize and reconstruct images.

3 Related Work

In recent years, dilated convolutions[5] have been proposed as a way to increase the receptive field of CNNs without increasing the number of parameters. Dilated convolutions, also known as atrous convolutions, involve inserting gaps between the kernel elements of a convolutional layer, effectively increasing the stride of the convolution operation. This allows the CNN to process larger regions of the input image while using the same number of parameters, thus increasing the effective receptive field.

While dilated convolutions have been shown to be effective in increasing the receptive field of CNNs, they still have limitations in capturing global information in the image. Specifically, dilated convolutions are still limited by the size of the input image, and the effective receptive field can only be increased up to a certain point before the model becomes too computationally expensive. In addition, dilated convolutions are not able to capture information that is spread across different frequency bands of the image, which is important for recognizing patterns and structures that occur at different scales.

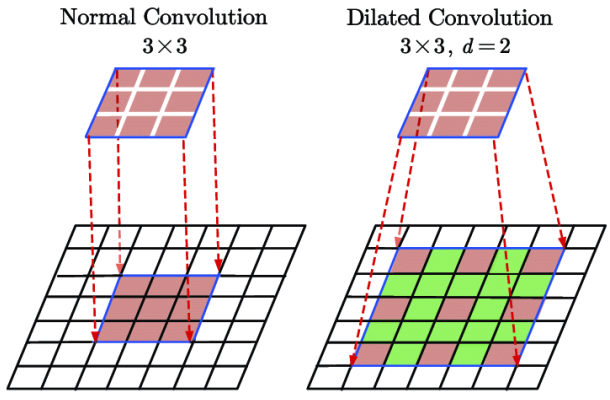

Figure 3: Receptive field for dilated conv

Another approach that has been proposed to increase the effective receptive field of CNNs is the use of attention mechanisms[6]. Attention mechanisms aim to selectively focus on important regions of the image, allowing CNN to capture global information by giving more weight to informative regions while ignoring less relevant ones.

While attention mechanisms have been shown to be effective in improving the performance of CNNs in various tasks, they still have limitations in capturing global information

in the image. Specifically, attention mechanisms are dependent on the spatial location of the feature maps and are not able to capture information that is spread across different frequency bands of the image. This makes them less effective in recognizing patterns and structures that occur at different scales, which can limit their performance in tasks that require capturing global information.

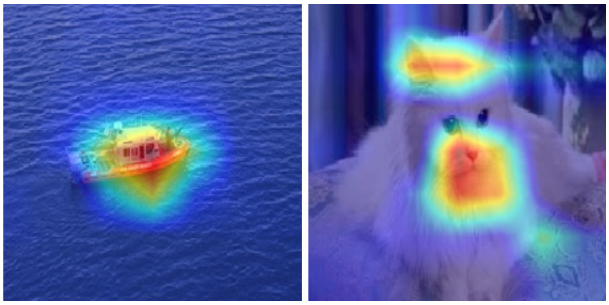

Figure 4: Reception field after attention

Fast Fourier Convolution (FFC)[7] is a recently developed frequency-based approach for image processing that replaces the traditional spatial convolution operation in CNNs with convolution in the frequency domain. FFC applies the Fast Fourier Transform (FFT) to the input image and then multiplies them in the frequency domain to obtain the output feature maps. This approach allows for the efficient computation of the convolution operation and also provides the ability to capture global information across different scales.

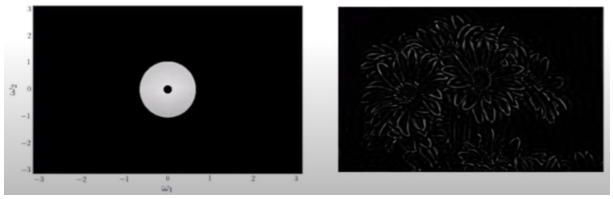

Figure 5: Only high frequencies of image(edges)

To the best of our knowledge, there is no follow-up work that uses wavelet instead of Fourier in FFC. The wavelet transform is a technique that decomposes a signal into different frequency components, while also capturing the temporal details of the signal, making it more useful than the Fourier transform which only captures frequency information. The wavelet transform is particularly well-suited for non-stationary signals and provides good frequency resolution for low-frequency components and good temporal resolution for high-frequency components, overcoming the limitations of the Fourier transform due to Heisenberg's Uncertainty Principle[8].

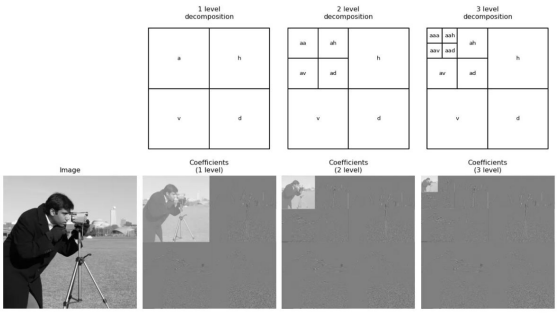

Figure 6: Wavelet decomposition

4 Frequency for receptive field

4.1 Mathematical perspective

From a linear algebra perspective, the grayscale image of size h by w is a vector of the discrete value of function I and is an element of R^{hxw} space. The Fourier transform's aim is to convert the vector from the standard basis of this space to the basis of Fourier space $e^{2\pi ijk/n}$, where j=0:hw. New coordinates of the image in Fourier space would be $\hat{I}_k = \sum_{j=0}^{k*w-1} I_j e^{-i2\pi jk/n}$ and $\hat{I}_k$ is exactly the length of orthogonal projection onto new basis vector. The simplest analog for this is the matrix diagonalization - we can find the basis (eigenvectors) where the matrix became diagonal. In geometric meaning, the ellipse transforms into a circle on a new basis. So, the image in Fourier space also has some useful properties that are helpful for feature extraction. For example, in figure 2 we see that high frequencies contain information about the edges of the image, which is hard to extract from raw pixels due to scale and intensity changes. Below is a more detailed review of each frequency transformation - Fourier and wavelet.

4.2 Fourier transform

4.2.1 Fourier series and Fourier transform

A lot of mathematical and computational problems can be solved by changing the coordinate system to the one where expressions simplify and are amenable to analysis. One of such transformations was introduced by Fourier. He observed that the sine and cosine functions of increasing frequency provide an orthogonal basis for the space of functions: the sines and cosines can be thought of as eigenvectors and frequencies - as eigenvalues.

A fundamental result in Fourier analysis is that a periodic and piece-wise smooth function $f(x)$ can be written as an infinite sum of sines and cosines of increasing frequency. For the L-periodic function $f(x)$ on $[-L, L)$ holds:

$$f(x) = \frac{a_0}{2} + \Sigma_{k=1}^{\infty}(a_k \cos(\frac{\pi kx}{L}) + b_k \sin(\frac{\pi kx}{L})),$$

with the coefficients

$$a_k = \frac{1}{2L}\int_{-L}^{L} f(x)\cos(\frac{\pi kx}{L})dx, b_k = \frac{1}{2L}\int_{-L}^{L} f(x)\sin(\frac{\pi kx}{L})dx.$$

It is more natural to write Fourier series in complex form using the Euler's formula:

$$f(x) = \Sigma_{k=-\infty}^{\infty}(a_k + ib_k)(\cos(\frac{\pi kx}{L}) + i\sin(\frac{\pi kx}{L})) = \Sigma_{k=-\infty}^{\infty} c_k e^{\frac{i\pi kx}{L}}$$

with the coefficients

$$c_k = \frac{1}{2L}\langle f(x), e^{\frac{-ik\pi x}{L}}\rangle = \frac{1}{2L}\int_{-L}^{L} f(x)e^{\frac{-ik\pi x}{L}}dx.$$

That is, $f(x)$ is represented as a sum of sines and cosines with a discrete set of frequencies $\omega_k = \frac{k\pi}{L}$, which becomes a continuous range as $L \to \infty$. Define $\omega = \frac{k\pi}{L}, \Delta\omega = \frac{\pi}{L}$ and take the limit as $L \to \infty$:

$$f(x) = \lim_{\Delta\omega \to 0} \Sigma_{k=-\infty}^{\infty} \frac{\Delta\omega}{2\pi} \int_{\frac{-\pi}{\Delta\omega}}^{\frac{\pi}{\Delta\omega}} f(\xi)e^{-ik\Delta\omega\xi}d\xi e^{ik\Delta\omega x}.$$

After taking the limit,

$$\int_{\frac{-\pi}{\Delta\omega}}^{\frac{\pi}{\Delta\omega}} f(\xi)e^{-ik\Delta\omega\xi}d\xi = \langle f(x), e^{-ikx}\rangle$$

becomes the Fourier transform of $f(x)$ denoted by $\hat{f}(\omega) := \mathcal{F}(f(x))$. The summation becomes a Riemann integral, giving the Fourier transform pair:

$$f(x) = (F)^{-1}(\hat{f}(\omega)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega x}d\omega,$$

$$\hat{f}(\omega) = \mathcal{F}(f(x)) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}dx.$$

4.2.2 Discrete Fourier transform

The above definitions of Fourier series and transform were stated for continuous functions. However, in the real day problems, it often makes more sense to consider vectors $\mathbf{f} = [f_0, ..., f_{n-1}]^T$ obtained by discretizing $f(x)$ at a regular spacing δx. According to this, the discrete Fourier transform (DTF) and its inverse (iDTF) are defined as:

$$\hat{f}_k = \Sigma_{j=0}^{n-1} f_j e^{\frac{-i2\pi jk}{n}},$$

$$f_k = \Sigma_{j=0}^{n-1} \hat{f}_j e^{\frac{i2\pi jk}{n}}.$$

That is, the DTF is a linear operator $\{f_0, ..., f_{n-1}\} \to \{\hat{f}_0, ..., \hat{f}_{n-1}\}$ and which with the notation $\omega_n = e^{\frac{-2\pi i}{n}}$:

$$\begin{pmatrix} \hat{f}_0 \\ \hat{f}_1 \\ . \\ . \\ . \\ \hat{f}_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & ... & 1 \\ 1 & \omega_n & \omega_n^2 & ... & \omega_n^{n-1} \\ 1 & \omega_n^2 & \omega_n^4 & ... & \omega_n^{2(n-1)} \\ & ... & & & \\ 1 & \omega_n^{n-1} & \omega_n^{2(n-1)} & ... & \omega_n^{(n-1)^2} \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ . \\ . \\ . \\ f_{n-1} \end{pmatrix}.$$

4.2.3 Fast Fourier transform

The above matrix multiplication is not very efficient and requires $\mathcal{O}(n^2)$ operations. This complexity can be improved by using the Fast Fourier transform. The main idea of all FFT methods is based on using the symmetry of the matrix and the periodicity of its elements. In particular, for any k holds:

$$\hat{f}_{n+k} = \Sigma_{j=0}^{n-1} f_j(x) e^{\frac{-i2\pi(n+k)j}{n}} = \Sigma_{j=0}^{n-1} f_j(x) e^{-i2\pi j} e^{\frac{-i2\pi kj}{n}} = \Sigma_{j=0}^{n-1} f_j(x) e^{\frac{-i2\pi kj}{n}},$$

which means that

$$\hat{f}_{n+k} = \hat{f}_k.$$

The basic idea behind the FFT is that the DFT may be implemented much more efficiently if the number of data points n is a power of 2. In case n is not a power of 2, it is still efficient to add additional zeroes to form an appropriate number. The idea is to divide the DFT computation into smaller parts:

$$\hat{f}_k = \Sigma_{j=0}^{n-1} f_j(x) e^{\frac{-i2\pi kj}{n}} =$$

$$= \Sigma_{m=0}^{\frac{n}{2}-1} f_{2m}(x) e^{\frac{-i2\pi k(2m)}{n}} + \Sigma_{m=0}^{\frac{n}{2}-1} f_{2m+1}(x) e^{\frac{-i2\pi k(2m+1)}{n}} =$$

$$= \Sigma_{m=0}^{\frac{n}{2}-1} f_{2m}(x) e^{\frac{-i2\pi k(m)}{n/2}} + e^{\frac{-i2\pi k}{n}} \Sigma_{m=0}^{\frac{n}{2}-1} f_{2m+1}(x) e^{\frac{-2\pi jm}{n/2}}.$$

That is, the initial DTF splits into two parts which are DTFs themselves, on the even and odd-numbered values. Because the above-mentioned symmetry, the number of needed calculations reduces. Reapplying this divide-and-conquer approach provides the computational complexity $\mathcal{O}(n \log n)$.

4.2.4 Pseudocode

In the following pseudocode for FFT we assume that A is a vector of length n, n is a power of 2 and ω is the n-th root of 1.

function FFT(A, ω**)**

Input: Coefficient representation of a polynomial $A(x)$ *of degree* $\leq n-1$*, where* n *is a power of 2.*

Output: Value representation $A(\omega^0), ..., A(\omega^{n-1})$

if $\omega = 1$: return $A(1)$

express $A(x)$ in the form $Aeven(x^2) + xAodd(x^2)$

call FFT($Aeven, \omega^2$) to evaluate $Aeven$ at even powers of ω

call FFT($Aodd, \omega^2$) to evaluate $Aodd$ at odd powers of ω

for $j = 0$ to $n-1$:

compute $A(\omega^j) = Aeven(\omega^{2j}) + \omega^j Aodd(\omega^{2j})$

return $A(\omega^0), ..., A(\omega^{n-1})$.

4.2.5 Application to images

The Fourier transform is an important image-processing tool. It transforms the image from the spatial domain into the frequency domain equivalent; that is, from pixels representation to frequency one.

For the square image of size $n \times n$, the two-dimensional DFT is given by:

$$\hat{f}(k,l) = \Sigma_{i=0}^{n-1} \Sigma_{j=0}^{n-1} f(i,j) e^{-i2\pi(\frac{ki}{n} + \frac{lj}{n})},$$

where $f(a,b)$ is the image in the spatial domain, and the exponential part represents the basis function corresponding to $\hat{f}(k,l)$. The basis functions are sines and cosines waves with increasing frequencies: $\hat{f}(0,0)$ corresponds to the mean brightness and $\hat{f}(n-1,n-1)$ - to the highest frequency.

To re-transform the image back to the spatial domain, the inverse DFT is used:

$$f(a,b) = \frac{1}{n^2} \Sigma_{k=0}^{n-1} \Sigma_{l=0}^{n-1} \hat{f}(k,l) e^{i2\pi(\frac{ka}{n} + \frac{lb}{n})}.$$

The resulting image of the Fourier transform is complex number valued and can be displayed as two images: the real and imaginary parts, or a phase and magnitude. The most commonly used is the magnitude part, as it contains the most information on the geometric structure of the spatial domain image. However, to invert the image to the spacial domain after processing, both parts are needed.

4.3 Wavelet transform

4.3.1 Idea and motivation

One of the main limitations of the Fourier transform is that it doesn't provide temporal localization: the base functions oscillate on the whole axis. Thus, if an original function has some properties which are localized, then it might be hard for Fourier transform to catch these changes.

One of the ideas how to address this issue is to localize the base functions in space. This is the core idea of wavelet transform: to use a wave-like oscillating function that has zero mean is 0 outside of the bounded segment. Then, this wavelet function can be shifted (to fit a target function in a localized segment) and scaled (to fit different frequency properties of a target function).

Let $\psi(t)$ be the wavelet function. The following operations are performed to define a full set of functions:

- Shifting is performed by $\psi(t-k)$ which corresponds to shifting nonzero part of wavelet function by k to the right.
- Scaling - $\psi(\frac{t}{k})$ is scaled wavelet function: if $k > 1$ then it corresponds to lower frequency, and if $k \in [0,1)$ then it corresponds to higher frequency.

Then, by applying shift and scale with proper parameters, the wavelet can be transformed to mimic the localized part of target function.

4.3.2 Wavelet functions

There are many options how to select the base wavelet function $\psi(t)$.

Examples and the names of corresponding wavelet functions are displayed on figure 7.

The choice of wavelet function depends on the nature of the target signal.

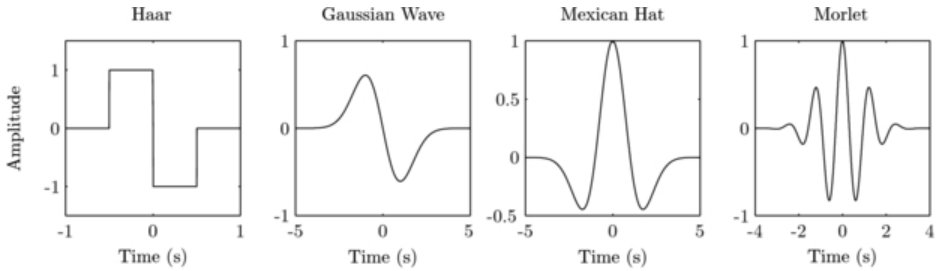

Figure 7: Examples of wavelet functions

4.3.3 Continious wavelet transform

Let $\psi(x)$ be the base wavelet function. The orthonormal basis for L^2 space is constructed from the set of functions $\psi_{ij}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)$: shift scaling is applied to base wavelet function to satisfy orthonormality of generated basis functions.

Then, the target function $f(x)$ can be represented as follows:

$$f(x) = \sum_{j,k=-\infty}^{j,j=+\infty} c_{jk}\psi_{jk}(x),$$

where c_{jk} are wavelet coefficients which can be determined based on wavelet transform:

$$[W_\psi f](a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} \psi(\tfrac{x-b}{a}) f(x)dx;$$

$$c_{jk} = [W_\psi f](2^{-j}, k2^{-j}).$$

4.3.4 Application to images

When wavelet transform is applied to a 2D image, it produces a decomposition into 4 components: a compressed version of an input image and its horizontal, vertical, and diagonal frequency components.

Wavelet transform is usually applied using both high-pass and low-pass filters, which are responsible for extracting edges or sharp components, and smooth image details correspondingly. These 2 filters are applied to an input image and then are recursively applied to each resulting component to the desired depth.

One level of the image wavelet transform scheme is displayed in the image below.

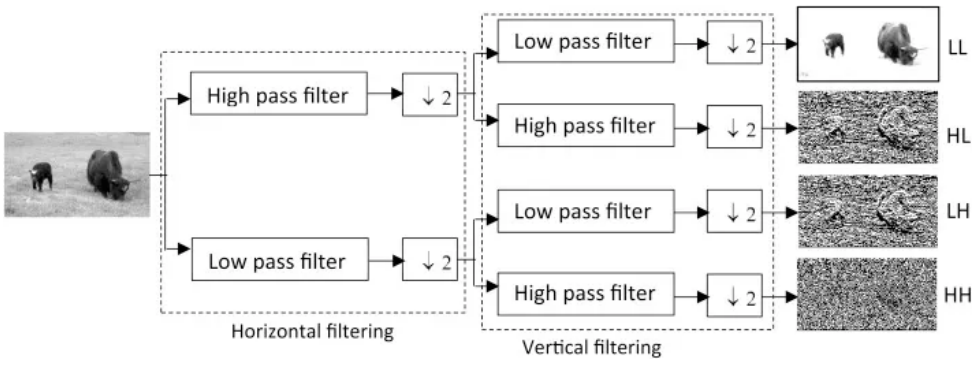

Figure 8: Wavelet transform application to the image

5 Experiments

For classification, we chose the cifar-10 dataset which contained images of 10 classes, 50k train images, and 10k images for testing. We train simple ResNet-like architecture on this dataset and compare it with injected frequency transformation for features in deep layers. We trained 3 models, first without frequency transformation, with Fourier transformation, and last with wavelet transformation. Below you can see the plot of validation accuracy on cifar-10. For the test accuracy, we have no transformation 67%, Fourier 70%, wavelet 73%.

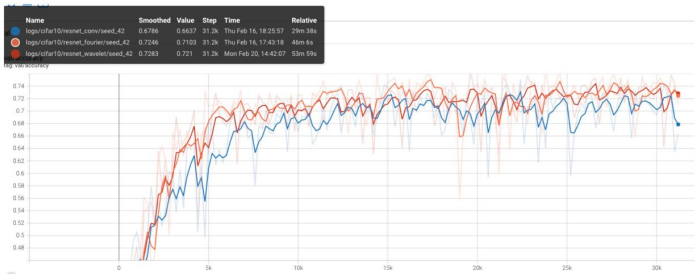

Figure 9: Validation accuracy, white for no transformation model, red for Fourier, green for wavelet

And lastly, for inpainting, we use an already trained model, because it requires much more resources to train. We compare the results of the model with dilated convolutions inside, regular, and with Fourier transformation.

Figure 10: Input image, regular result, dilated and Fourier results

From the images above it's clear that the regular model sees only local information and thus fails to inpainted region. The model with dilated conv has a bigger receptive field and can better restore, but still fails to correctly identify high and low frequencies and the model with Fourier gives the best result due to the non-local receptive field.

6 Conclusion

In conclusion, we explored the use of Fourier and wavelet transforms to improve the accuracy of Convolutional Neural Networks (CNNs) in image classification and inpainting tasks. By decomposing the image into its frequency components, we were able to capture global information and improve CNN's ability to recognize and reconstruct images. We compared the approach with each other: regular convolution, Fourier, and wavelet, and found that our approach outperformed regular convolution in capturing global information. Fast Fourier Convolution (FFC) is included in the traditional spatial convolution operation with convolution in the frequency domain, making it more accurate. Overall, our results demonstrate the potential of frequency transformations in improving the performance of CNNs in image processing tasks.

7 References

[1] An Introduction to Convolutional Neural Networks. Keiron O'Shea, Ryan Nash.
[2] Advancements in Image Classification using Convolutional Neural Network
[3] Image Inpainting for High-Resolution Textures using CNN Texture Synthesis. Pascal Laube, Michael Grunwald, Matthias O. Franz, Georg Umlauf
[4] Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. Wenjie Luo, Yujia Li, Raquel Urtasun, Richard Zemel
[5] Multi-Scale Context Aggregation by Dilated Convolutions. Fisher Yu, Vladlen Koltun
[6] Attention operates uniformly throughout the classical receptive field and the surround. Bram-Ernst Verhoef Is a corresponding author, John HR Maunsell
[7] Fast Fourier Convolution. Lu Chi, Borui Jiang, Yadong Mu
[8] Sen, D. (2014). "The Uncertainty relations in quantum mechanics". Current Science. 107 (2): 203–218.