Ukrainian Catholic University

Faculty of Applied Sciences

Data Science Master Programme

Frequency transformation for image recognition and inpainting

Linear Algebra intermediate report

Authors:
Igor Babin, Kostiantyn Verhun, Iryna Pastukhova

22 January 2023

1 A clear statement of the topic and the aim of your project

Improve the receptive field of Convolution Neural Networks(CNN) using frequency transformation(Fourier and wavelet) and show it's an advantage in image inpainting.

2 Description of the problem you will discuss

The classical CNN approach for tasks like image classification is to stack convolution filters, which are then trained to detect some features on images. Lower layers learn to predict basic information, like edges, and upper layers learn to detect semantic information, like parts of the body. The limitation of this method is that the convolution filter can see only a limited region of image on each step, called the receptive field Figure 1

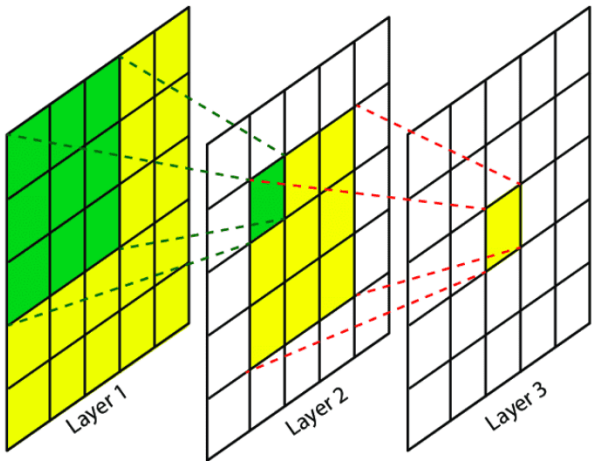

Figure 1: Receptive field for 3 layers of CNN

Image inpainting is the task of restoring some damaged or masked region. Most existing DL approaches use Image-to-Image fully convolutional networks, and having a high receptive field is essential.

Figure 2: Example of image inpainting

3 Sketch of the possible approaches

The first approach to solve this issue is to use dilated convolution with no neighbor pixels but with some steps. This solution helps but still fails to capture all information in one.

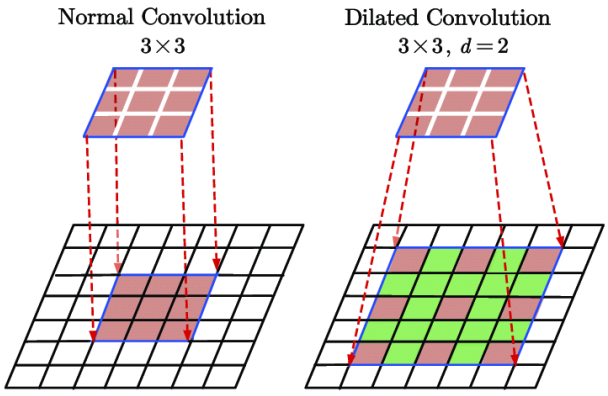

Figure 3: Receptive field for dilated conv

Another approach is to use an attention mechanism, which directly compares all pixel in pairs and construct an attention map that highlights important regions. An issue of this approach it's high computation complexity $O(n^2)$ and requires a lot of data to train.

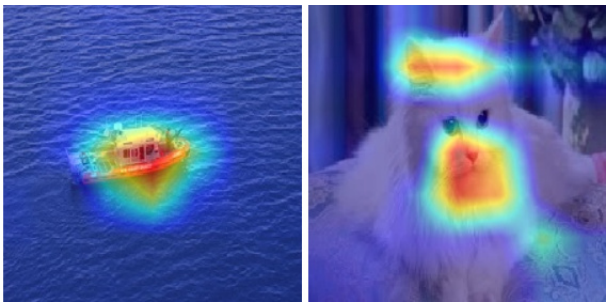

Figure 4: Reception field after attention

The novel approach to using Fourier transformation on the image (or features), where the frequency is a change between neighboring pixels, so each Fourier coefficient will contain information about the whole image in the frequency domain. Thus convolution will see all information in some frequency globally. A possible extension of Fourier transformation is to use Wavelet transformation because it increases temporal information in the frequency domain, making it not global but semi-global.

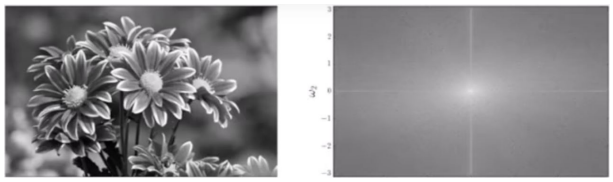

Figure 5: Fourier transformation of image

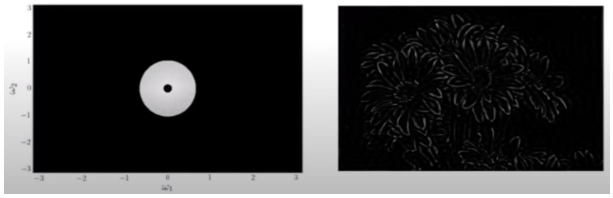

Figure 6: Only high frequencies of image(edges)

4 Short explanations of the pros and cons of your chosen method

The pros of the method are a high receptive field and, as a result, better generalization to higher dimensions, low calculation complexity $O(NlogN)$. Cons - higher computational complexity compared to vanilla convolution.

5 A short discussion of the planned research steps and/or numerical realization

We'll use a simple classification model on raw pixels and frequencies to test receptive field improvements.

To clarify the importance of the Fourier block on image inpainting, we'll compare the results of the pre-trained network with and without this block.