Progress Report

For

Spectral Leakage and Rethinking the Kernel Size in CNNs

At

https://openaccess.thecvf.com/content/ICCV2021/papers/Tomen_Spectral_Leakage_and_Rethinking_the_Kernel_Size_in_CNNs_ICCV_2021_paper.pdf

Spectral Leakage -- appears when the number of periods of signal passed to discrete Fourier transform is not an integer. When applying a window function to the periodic signal, there is a chance that the cut window is not periodic and the sharp edges (sinc like function) will generate a bunch of frequencies that distorts the filtered image around the signal.

There are some approaches that try to solve this problem. In the article provided in the following link (https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8613772), a sampling operator composed of windowing and sampling operations is used to encode the sparse frequency representation of a filter, and use it to reconstruct the window function that depends on the size of every object, which suppresses the leakage problem. In the selected paper, instead of a dynamic window function, what we will reconstruct is going to be a fixed size kernel with a hamming window that helps tackle the spectral leakage issue.

In this paper, the authors consider the well-known problem of spectral leakage caused by windowing artifacts in filtering operations in the context of CNNs, and they showed that by reducing the spectral leakage, the classification accuracy of the CNN would be increased, and its robustness would also be improved. Their focus is about improving the accuracy of CNNs with large kernels by using the Hamming window function as small size kernels (k = 3) would make them susceptible to spectral leakage. They demonstrated their result on Fashion-MNIST, CIFAR-10, CIFAR-100 and ImageNet with the simple use of a standard window function in convolutional layers and showed that CNNs employing the Hamming window display increased robustness against various adversarial attacks.

At the beginning of the paper, authors showed that a standard 7x7 CNN kernel trained on CIFAR-10 struggles to learn good quality bandpass filters, as the use of small kernel sizes typically lead to severe truncation, an example would be Gabor filter with severe truncation would lead to spectral leakage in the frequency response due to sinc artifacts, while the same filter with negligible truncation (kernel_size being 49 x 49) is a good bandpass filter. And the way to fix this, proposed by the authors, is by the use of a standard Hamming window to taper off the kernels in the space domain.

Since the use of smaller kernel size can help reduce the computational complexity and improves accuracy, it has become more of an unspoken rule for people to use small sized kernels for their CNNs. And although larger kernels may cause over-parameterization, the use of windowing method can effectively constrain the parameter space, and it's not like the use of weight decay, dropout, early stopping, and data augmentation, this windowing method can encourage a center-bias in the kernel shape, and it works well with the use of weight decay and data augmentation. The Hamming window can be interpreted as a form of regularization. And the author's approach is done by performing simple multiplication in the space domain.

For CNNs, when given as input an adversarial image with perturbations tiny in magnitude, it could produce a wrong classification result with high confidence, while for humans, the perturbation is barely noticeable. Hence, spectral leakage artifacts would cause

trouble for the classification process as the artifacts are typically small in magnitude, but they may be present in every feature map in a standard CNN with non-tapered filters.

So, instead of using the small sized kernels, the authors used the standard Hamming window in order to reduce unwanted frequency components, and the Hamming window can be implemented by multiplying each 2-dimensional k x k kernel in a convolutional layer with the k x k Hamming window function.

The author also did a fully controlled experiment to test whether the kernels in a single convolutional layer trained in a supervised setting display spectral leakage; they forced the network to learn good quality bandpass filters in a regression task to predict the FFT magnitude of the input image. For this experiment, they designed 2 different CNNs, one with a Hamming window, and one without. And as a result, they found that the model using Hamming windows alleviates leakage artifacts. Also, by visualising the predictions of the 2 CNNs, it's obvious that the bandpass filters learned by the CNN without Hamming window do suffer from leakage artifacts, and a Hamming window regularizes the kernel weights and fights against it.

When performing their CNN on the CIFAR-10 dataset, the kernel size was 7x7, and the use of Hamming window in all convolutional layers caused a significant boost to the validation accuracy. And by doing another controlled experiment, they also concluded that the performance increase that the Hamming window provided is independent of aliasing and the choice of downsampling method. They also found that in explicitly regularized networks, the accuracy boost caused by their windowing method got ever larger. They also did a comparison between a normal CNN and their CNN, while using kernel size k = 3 for normal CNN, and kernel size = 7x7 and 9x9 for their CNN, they found that their CNN (doesn't matter k = 7x7 or k = 9x9) got much better accuracy result compared with the normal CNN with kernel size k = 3. Authors also tried Hann and Blackman window instead of the Hamming window, and found that they provide the same performance boost to CNN as the Hamming window does. Authors also did similar experiments on CIFAR-100, and found that the windowed network still had better performance than normal CNN constantly.

They then tried their network on Fashion-MNIST and MNIST datasets, which contain less natural images, where not all frequency components are well-represented in the training set, and the effects of windowing would be less-prominent. With Fashion-MNIST, the authors found similar results with the network performing on CIFAR-10 and CIFAR-100, their network would have better performance than normal CNNs, but for MNIST, they didn't find performance increase for their windowed network. For this, they thought it was because the lack of high frequency components for images in MNIST dataset as leakage in lowpass and bandpass filters cannot contaminate high frequency information. So, they subsampled each image in MNIST from 28x28 to 14x14 which increased the relative magnitude of the high frequency component and in this way, their windowed networks again achieved much better classification results than normal CNNs. As for the ImageNet dataset, they trained another windowed CNN with kernel size k = 7, and found that their network got better performance.

Finally, the authors tested the robustness of their windowed CNN on CIFAR-10 against DeepFool, and found that although for network with kernel size k = 5, the Hamming model didn't performance as good as baseline CNN, but for large kernel sizes, the robustness of Hamming models is significantly better than baseline CNNs.