Uses of Complex Wavelets in Deep Convolutional Neural Networks

Fergal Brian Cotter

August 15, 2019

Image understanding has long been a goal for computer vision. It has proved to be an exceptionally difficult task due to the large amounts of variability that are inherent to objects in a scene. Recent advances in supervised learning methods, particularly convolutional neural networks (CNNs), have pushed forth the frontier of what we have been able to train computers to do.

Despite their successes, the mechanics of how these networks are able to recognize objects are little understood, and the networks themselves are often very difficult and time-consuming to train. It is very important that we improve our current approaches in every way possible.

A CNN is built from connecting many learned convolutional layers in series. These convolutional layers are fairly crude in terms of signal processing - they are arbitrary taps of a finite impulse response filter, learned through stochastic gradient descent from random initial conditions. We believe that if we reformulate the problem, we may achieve many insights and benefits in training CNNs. Noting that modern CNNs are mostly viewed from and analyzed in the spatial domain, this thesis aims to view the convolutional layers in the frequency domain (viewing things in the frequency domain has proved useful in the past for denoising, filter design, compression and many other tasks). In particular, we use *complex wavelets* (rather than the Fourier transform or the discrete wavelet transform) as basis functions to reformulate image understanding with deep networks.

In this thesis, we explore the most popular and well-developed form of using complex wavelets in deep learning, the ScatterNet from Stephane Mallat. We explore its current limitations by building a DeScatterNet and found that while it has many nice properties, it may not be sensitive to the most appropriate shapes for understanding natural images.

We then develop a *locally invariant* convolutional layer, a combination of a complex wavelet transform, a modulus operation, and a learned mixing. To do this, we derive backpropagation equations and allow gradients to flow back through the (previously fixed) ScatterNet front end. Connecting several such locally invariant layers allows us to build *learnable ScatterNet*, a more flexible and general form of the ScatterNet (while still maintaining its desired properties).

We show that the learnable ScatterNet can provide significant improvements

over the regular ScatterNet when being used as a front end for a learning system. Additionally, we show that the locally invariant convolutional layer can directly replace convolutional layers in a deep CNN (and not just at the front-end). The locally invariant convolutional layers naturally downsample the input (because of the complex modulus) while increasing the channel dimension (because of the multiple wavelet orientations used). This is an operation that often happens in a CNN by a combination of a pooling and convolutional layer. It was at these locations in a CNN where the learnable ScatterNet performed best, implying it may be useful as learnable pooling layer.

Finally, we develop a system to learn complex weights that act directly on the wavelet coefficients of signals, in place of a convolutional layer. We call this layer the *wavelet gain layer* and show it can be used alongside convolutional layers. The network designer may then choose to learn in the pixel *or* wavelet domains. This layer shows a lot of promise and affords more control over what regions of the frequency space we want our layer to learn from. Our experiments show that it can improve on learning in the pixel domain for early layers of a CNN.