Complex Valued Convolutional Neural Networks for Image Restoration

Amara Datta Dola (19MA20003)

Thesis Mid-Semester

Supervised by:

Prof. Mahesh Mohan M R and
Prof. Dipankar Ghosh

Department of Mathematics, IIT Kharagpur

December 4, 2023

Introduction

- **Convolutional Neural Networks (CNNs):** Existing literature in computer vision typically focus on real-valued data.
- **Merits of Complex-Valued Signals:** Having diverse applications in SAR, MRI, and meteorology.
- **Novel CNN Variation:** Introduction of CNNs utilizing complex-valued inputs and weights, enhancing phase capture in images.
- **Challenges and Exploration:** Adapting optimization techniques, like backpropagation, for complex-valued neural networks.

Motivation

- **Noise Reduction via FFT:**
  Starting with a noisy image, FFT translates it into frequency domain. A circular low-pass filter removes high-frequency noise, and Inverse FFT restores the image with reduced noise and enhanced clarity.

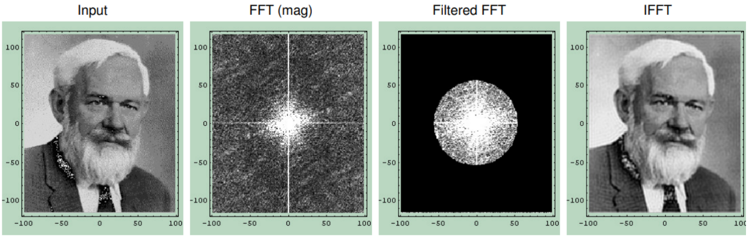

Figure 1: Process of Image Restoration using Fourier transform

Orthogonal Decision Boundaries

- Complex-valued neurons create decision boundaries consisting of two intersecting hypersurfaces at orthogonal angles, enhancing classification capabilities.
- Single complex-valued neuron with n-inputs is equivalent to two real-valued neurons with 2n-inputs which have a restriction on a set of weight parameters.
- For a single neuron with a single input,complex weight $W = w_1 + iw_2$. Let M be the matrix representing the transformation:

$$\mathbf{M} = \begin{pmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{pmatrix} = |W| \begin{pmatrix} \cos(\text{arg}(W)) & -\sin(\text{arg}(W)) \\ \sin(\text{arg}(W)) & \cos(\text{arg}(W)) \end{pmatrix}$$

- This orthogonality improves generalization. Example, several problems like XOR that cannot be solved with a single real neuron, can be solved with a single complex-valued neuron using the orthogonal property.

Other Advantages of using CVNNs

- **Complex Activation Functions**: Real-valued non-linear activations cannot maintain the magnitude and phase information of complex-valued inputs. Complex-valued activations in CVNNs preserve the intricate relationships embedded in complex data's magnitude and phase [1].
- **Enhanced Information Encoding:** Complex numbers allow for the encoding of more information in the same number of parameters.
- **Robustness to Noise:** Their ability to handle complex representations allows them to effectively discard and filter out irrelevant fluctuations, leading to more accurate and reliable outputs in noisy environments.

[1] ON COMPLEX-DOMAIN CNN REPRESENTATIONS FOR CLASSIFYING REAL/COMPLEX-VALUED DATA Mahesh Mohan M R, K Srivastava, N Ahuja, 2023 (under review)

Complex Calculus Preliminaries

Definition

A complex function f is complex differentiable at z, if the following limit exists

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}.$$

Definition

A complex function f is complex differentiable at point z iff u, v are differentiable (as real functions) , and the Cauchy-Riemann equations hold at z:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Definition

Liouville's Theorem:
If f(z) is an analytic function for all finite values of z and is bounded for all values of z in C, then f is a constant function.

Challenges in optimizing loss function

- Complex functions need to be holomorphic (analytic) and complex differentiable for gradient to exist, but many complex functions are not analytic.
- Issues using complex loss function:
  - If the complex loss function f is analytic and bounded everywhere, it is constant as implied by Liouville's Theorem.
  - Also, if the loss function is complex, there is no notion of orderedness in the complex loss function.
- If the loss function f is real-valued and analytic everywhere (i.e., $v = 0$), the Cauchy-Riemann (CR) equations simplify to $\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0$, implying that if f must be constant.
- **Solution: Wirtinger derivatives provide an alternative approach to complex derivatives.**

Introducing Wirtinger derivatives

Suppose $f(z) = u(x, y) + iv(x, y)$ and $h = h_1 + ih_2$.
Let's break down f(z+h) using first order Taylor's series:

$$u(c+h) = u(c) + \frac{\partial u}{\partial x}(c)h_1 + \frac{\partial u}{\partial y}(c)h_2 + o(|h|),$$
$$v(c+h) = v(c) + \frac{\partial v}{\partial x}(c)h_1 + \frac{\partial v}{\partial y}(c)h_2 + o(|h|).$$

Multiplying the second relation by i and adding it to the first one, we get:

$$f(c+h) = f(c) + \left(\frac{\partial u}{\partial x}(c) + i\frac{\partial v}{\partial x}(c)\right) h_1 + \left(\frac{\partial u}{\partial y}(c) + i\frac{\partial v}{\partial y}(c)\right) h_2 + o(|h|).$$

Next, we substitute h_1 and h_2 using the relations $h_1 = \frac{h+h^*}{2}$ and $h_2 = \frac{h-h^*}{2i}$.

$$f(c+h) = f(c) + \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) - i\frac{\partial f}{\partial y}(c)\right) h + \frac{1}{2}\left(\frac{\partial f}{\partial x}(c) + i\frac{\partial f}{\partial y}(c)\right) h^* + o(|h|).$$

Defining Wirtinger derivatives

$$f(c+h) = f(c) + \frac{1}{2}\left(\frac{\partial f}{\partial z}(c) - i\frac{\partial f}{\partial z^*}(c)\right)h + \frac{1}{2}\left(\frac{\partial f}{\partial z}(c) + i\frac{\partial f}{\partial z^*}(c)\right)h^* + o(|h|).$$

Definition

The Wirtinger derivative operators are

$$\frac{\partial}{\partial z} := \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)$$

$$\frac{\partial}{\partial z^*} := \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

In view of these new definitions, equation may now be recasted as follows:

$$f(c+h) = f(c) + \frac{\partial f}{\partial z}(c)h + \frac{\partial f}{\partial z^*}(c)h^* + o(|h|).$$

Properties of wirtinger derivatives

Definition

The total differential of any complex valued function f is

$$df = \frac{\partial f}{\partial z}dz + \frac{\partial f}{\partial z^*}dz^*$$

Theorem

Given two complex functions $f(z)$ and $g(z)$, and any two complex constants a and b, the linearity property of Wirtinger derivatives is expressed as:

1. *For the derivative with respect to z:*

$$\frac{\partial}{\partial z}(af(z) + bg(z)) = a\frac{\partial f}{\partial z} + b\frac{\partial g}{\partial z}$$

2. *For the derivative with respect to the conjugate of z (denoted as z^*):*

$$\frac{\partial}{\partial z^*}(af(z) + bg(z)) = a\frac{\partial f}{\partial z^*} + b\frac{\partial g}{\partial z^*}$$

Conjugate property of wirtinger derivative

Theorem

If u, v is differentiable at a point c, then the following equality holds:

$$\left(\frac{\partial f}{\partial z}(c)\right)^* = \frac{\partial f^*}{\partial z^*}(c). \tag{1}$$

Theorem

If u, v is differentiable at a point c, then the following equality holds:

$$\left(\frac{\partial f}{\partial z^*}(c)\right)^* = \frac{\partial f^*}{\partial z}(c). \tag{2}$$

This result complements the previous one, showing the symmetry in the behavior of the Wirtinger derivatives with respect to complex conjugation.

Other properties

Theorem

Product Rule:
If f, g are differentiable in the real sense at c, then

$$\frac{\partial(f \cdot g)}{\partial z}(c) = \frac{\partial f}{\partial z}(c)g(c) + f(c)\frac{\partial g}{\partial z}(c),$$
$$\frac{\partial(f \cdot g)}{\partial z^*}(c) = \frac{\partial f}{\partial z^*}(c)g(c) + f(c)\frac{\partial g}{\partial z^*}(c).$$

Sketch of the proof: Let $f(z) = f(x,y) = u_f(x,y) + iv_f(x,y), g(z) = g(x,y) = u_g(x,y) + iv_g(x,y)$ be two complex functions differentiable at c. Consider the complex function r defined as $r(z) = f(z)g(z)$. Then

$$r(z) = (u_f(z) + iv_g(z))(u_g(z) + iv_g(z)) = (u_f u_g - v_f v_g) + i(u_f v_g + v_f u_g).$$

We proved the product rule for the operators $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ and this proves the product rule for W and CW derivatives.

Other properties

Theorem

Reciprocal Rule: *If f is differentiable in c and $f(c) \neq 0$, then*

$$\frac{\partial\left(\frac{1}{f}\right)}{\partial z}(c) = -\frac{\frac{\partial f}{\partial z}(c)}{f^2(c)}, \quad \frac{\partial\left(\frac{1}{f}\right)}{\partial z^*}(c) = -\frac{\frac{\partial f}{\partial z^*}(c)}{f^2(c)}$$

Theorem

Division Rule: *If f, g are differentiable in the real sense at c and $g(c) \neq 0$, then*

$$\frac{\partial\left(\frac{f}{g}\right)}{\partial z}(c) = \frac{\frac{\partial f}{\partial z}(c)g(c) - f(c)\frac{\partial g}{\partial z}(c)}{g^2(c)},$$

$$\frac{\partial\left(\frac{f}{g}\right)}{\partial z}(c) = \frac{\frac{\partial f}{\partial z^*}(c)g(c) - f(c)\frac{\partial g}{\partial z^*}(c)}{g^2(c)}.$$

The division rule follows immediately from the multiplication rule and the reciprocal rule $\left(\frac{f(c)}{g(c)} = f(c) \cdot \frac{1}{g(c)}\right)$.

Chain Rule for Backpropagation

Theorem

If f, g are differentiable in the real sense i.e u_f, u_g, v_f, v_g are differentiable at c and $g(c) \neq 0$, then

$$\frac{\partial g \circ f}{\partial z}(c) = \frac{\partial g}{\partial z}(f(c))\frac{\partial f}{\partial z}(c) + \frac{\partial g}{\partial z^*}(f(c))\frac{\partial f^*}{\partial z}(c),$$
$$\frac{\partial g \circ f}{\partial z^*}(c) = \frac{\partial g}{\partial z}(f(c))\frac{\partial f}{\partial z^*}(c) + \frac{\partial g}{\partial z^*}(f(c))\frac{\partial f^*}{\partial z^*}(c).$$

Sketch of the Proof: Let's write the total differential for $f(g)$:

$$d(f \circ g) = \frac{\partial f}{\partial g}dg + \frac{\partial f}{\partial g^*}dg^*$$

Put dg and dg^* in to the equation:

$$d(f \circ g) = \left(\frac{\partial f}{dg}\frac{\partial g}{dz} + \frac{\partial f}{dg^*}\frac{\partial g^*}{dz}\right)dz + \left(\frac{\partial f}{dg}\frac{\partial g}{dz^*} + \frac{\partial f}{dg^*}\frac{\partial g^*}{dz^*}\right)dz^*$$

So, these are the chain rules and they are exactly same with the one we know for real functions! (as thinking $f(g, g^*), g(z, z^*)$ are real multi-variable functions)

Optimizing Single Variable Real Loss Function

Theorem

If f is a real valued function(loss function) defined on $\mathbb{C}$, the steepest descent's direction is $-\frac{\partial f}{\partial z^*}$

Its first order Taylor's expansion at z is given by:

$$f(c+h) = f(c) + \frac{\partial f}{\partial z}(c)h + \left(\frac{\partial f}{\partial z}(c)h\right)^* + o(|h|) \quad = f(c) + 2\Re\left[\frac{\partial f}{\partial z}(c)h\right] + o(|h|).$$

However, in view of the Cauchy Riemann inequality we have:

$$\Re\left[\frac{\partial f}{\partial z}(c)h\right] = \Re\left[\left\langle h, \left(\frac{\partial f}{\partial z}(c)\right)^*\right\rangle_{\mathbb{C}}\right] \leq \left|\left\langle h, \left(\frac{\partial f}{\partial z}(c)\right)^*\right\rangle_{\mathbb{C}}\right| \leq |h| \left|\frac{\partial f}{\partial z^*}(c)\right|.$$

The equality in the above relationship holds, if $h \uparrow \frac{\partial f}{\partial z^*}$. Direction of steepest descent is -$\frac{\partial f}{\partial z^*}$. Update scheme of gradient descent based algorithm minimizing $f(z)$ is:

$$z_n = z_{n-1} - \mu \cdot \frac{\partial f}{\partial z^*}\left(z_{n-1}\right), \mu > 0$$

Optimizing Multivariate Loss Functions

Consider a multivariate complex function $f : \mathbb{C}^n \rightarrow \mathbb{R}$ and let $Z = (z_1, z_2, \ldots, z_n)$ be a point in the domain of f.The total derivative of f at Z is:

$$df = \sum_{j=1}^{n} \left(\frac{\partial f}{\partial z_j} h_j + \frac{\partial f}{\partial {z_j}^*} {h_j}^* \right).$$

Since f is a real valued loss function, it must hold that:

$$df = 2\text{Re} \left(\sum_{j=1}^{n} \frac{\partial f}{\partial z_j} h_j \right)$$

We define the gradient of f at Z w.r.t the complex variables as $\nabla f(Z) = \left(\frac{\partial f}{\partial z_1}, \ldots, \frac{\partial f}{\partial z_n} \right)^T$, and its conjugate transpose as $\nabla f(Z)^* = \left(\frac{\partial f}{\partial {z_1}^*}, \ldots, \frac{\partial f}{\partial {z_n}^*} \right)^T$.

Theorem

For the steepest descent, the direction is $-\nabla f(Z_i)^* = -\left(\frac{\partial f}{\partial {z_1}^*}, \ldots, \frac{\partial f}{\partial {z_n}^*} \right)^T$

For Complex Valued NN, we do gradient descent as follows: $Z_{i+1} = Z_i - \lambda \nabla f(Z_i)^*$

Conclusion

- We have conducted a systematic study of optimizing complex-valued CNNs. We concluded that normal complex valued differentiation cannot be used for optimizing complex-valued neural networks; instead, we need to consider a different calculus called Wirtinger calculus
- We have derived useful properties of the Wirtinger calculus that is needed for backpropagation.
- The expressiveness of Wirtinger derivatives are wonderful. Apart from chain rule, every property with Wirtinger derivatives becomes same as that one learned in real calculus.

Thank You

Any Questions?