# Frequency learning for structured CNN filters with Gaussian fractional derivatives (Supplementary)

Nikhil Saldanha,
Silvia L. Pintea,
Jan C. van Gemert,
Nergis Tomen

Computer Vision Lab,
Delft University of Technology,
Delft, Netherlands

## 1 Fractional-order derivatives

This section provides a narrow introduction to fractional calculus as required for understanding our approach. Let us take a function $f(x)$ which we would like to differentiate with respect to the variable $x$. This can be written as follows:

$$\frac{d}{dx}f(x) = D[f(x)] = D_x f. \tag{1}$$

$\frac{d}{dx}$ is the derivative operator with respect to the variable $x$. This formulation is known as a first-order derivative of $f$ with respect to $x$. $D_x f$ is a short-hand for the same. We can successively differentiate the function $f(x)$ $n$ times to get an $n^{th}$ order derivative. Here we are only concerned with positive integer orders, i.e. $n \in \mathbb{Z}^+$:

$$\frac{d^n}{dx^n}f(x) = D_x^n f. \tag{2}$$

The simplest definition of the fractional derivative operator would be one that can operate on integers as well as fractions. Let us use $\nu$ to represent real numbers that can take on integer or fractional values. Hence, in order for the fractional derivative operator $D^\nu(\cdot)$, $\nu \in \mathbb{R}$ to be properly defined, it should be accurate for integer values as well. One of the best and simplest tests for this is that the composition of derivatives of two fractional orders $\nu_1$ and $\nu_2$ should be equivalent to the derivative of the sum of the orders $\nu_1 + \nu_2$ of the composing derivatives:

$$D^{\nu_1}(D^{\nu_2} f) = D^{\nu_1+\nu_2} f. \tag{3}$$

In principle, fractional derivatives should be able to define normal derivatives $Df$ and also integrals $\int f$:

$$D^{\frac{1}{2}}(D^{\frac{1}{2}} f) = Df \tag{4}$$

$$D^{-1} f(x) = \int f(x)dx. \tag{5}$$

Note that there is no single form of a fractional derivative and much like integrals or even regular differentials, there are different ways of defining fractional derivatives for different types of functions. In the next section, we review fractional derivatives of some common functions.

## 1.1 Fractional derivatives of polynomial functions

We can first break up polynomials as terms that look like $ax^k$ and then each term can be differentiated separately using the chain rule of differentiation:

$$Dax^k = akx^{k-1} \tag{6}$$

$$D^2ax^k = ak(k-1)x^{k-2} \tag{7}$$

$$\vdots$$

$$D^n ax^k = a\frac{k!}{(k-n)!}x^{k-n}, k \geq n. \tag{8}$$

When considering non-integer values of the order $\nu$, the factorial function does not have solutions for all real numbers. To get around this issue, we can use the Gamma function $\Gamma(\cdot)$ which accepts as input all positive real numbers:

$$D^\nu ax^k = a\frac{\Gamma(k)}{\Gamma(k-\nu)}x^{k-\nu}. \tag{9}$$

## 1.2 Fractional derivatives of exponential functions

We can expand exponentials as an infinite series of polynomials and apply the same technique as in the previous section:

$$D^\nu e^x = D^\nu \sum_{i=0}^{\infty} \frac{x^i}{i!} = \sum_{i=0}^{\infty} \frac{x^{i-\nu}}{(i-\nu)!}, \tag{10}$$

where again for non-integer values of $\nu$, we use the Gamma function $\Gamma(\cdot)$:

$$D^\nu e^x = \sum_{i=0}^{\infty} \frac{x^{i-\nu}}{\Gamma(i-\nu)}. \tag{11}$$

We can combine the definitions of fractional order derivatives of polynomials and exponentials to define the fractional order derivative for the Gaussian family of functions.

## 1.3 Fractional-order derivative of Gaussians

### 1.3.1 Using polynomial expansion

We first express the Gaussian as a Taylor Series (about $x = 0$) to get the polynomial terms:

$$\begin{aligned} G(x;\sigma=1,\mu=0) &= \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}} - \frac{x^2}{2\sqrt{2\pi}} + \frac{x^4}{8\sqrt{2\pi}} - \frac{x^6}{48\sqrt{2\pi}} + \cdots \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k k!} \cdot \frac{x^{2k}}{\sqrt{2\pi}} \end{aligned} \tag{12}$$

Each term is now a polynomial term whose fractional order derivative can be described in terms of equation 9:

$$D^{\nu} G(x;\sigma=1,\mu=0) = \sum_{k=0}^{\infty} \frac{-1^k}{2^k k!} \cdot \frac{\Gamma(2k+1)}{\Gamma(2k-\nu+1)} \cdot \frac{x^{2k-\nu}}{\sqrt{2\pi}}. \tag{13}$$

The drawback of this approach is that it yields complex values for $x < 0$. This works perfectly fine at integer orders but does not give correct values for negative $x$ for fractional orders, since there we cannot deal with the imaginary component. One possible solution is to shift the Gaussian in the positive direction so we do not encounter negative values. The resulting Gaussian derivative should be identical in shape to the one at $x = 0$ since the mean of a Gaussian only determines its shape. By expanding the series about some point $x = a$, we get the following expression:

$$\begin{aligned} G(x;\sigma=1,\mu=0) = &\frac{ae^{-a^2/2}(x-a)}{\sqrt{2\pi}} + \frac{(a^2-1)e^{-a^2/2}(x-a)^2}{2\sqrt{2\pi}} \\ &- \frac{(a(a^2-3)e^{a^2/2})(x-a)^3}{6\sqrt{2\pi}} + \cdots \end{aligned} \tag{14}$$

Unfortunately, there is no closed-form expression for this expansion which is compounded by the fact that each of the terms of the expansion contains an exponential in the form of the Gaussian which if expanded again will generate more exponential terms. Hence we abandon this approach entirely and consider next the Caputo-Fabrizio form of the fractional derivative.

### 1.3.2 Using Caputo-Fabrizio closed-form

One of the more popular forms of the fractional derivative is called Caputo-Fabrizio (CF) fractional derivative [3, 4, 6]. We are interested in this form since prior work has shown to be able to capture multiple types of functions. One of the advantages of this form is that it allows us to separate the fractional part of the derivative from the integer order derivative and compose them as required. It has also been shown in previous work that a closed-form expression for the Gaussian derivative is possible [1]. The final closed-form expression looks as follows:

$${}^{CF}D_x^{n+\nu} G(x) = \frac{1}{1-\nu} \sum_{k=1}^{n} \left[ \left(-\frac{\nu}{1-\nu}\right)^{n-k} \partial_x^k G(x) \right] + \left(-\frac{\nu}{1-\nu}\right)^{n} {}^{CF}D_x^{\nu} G(x) \tag{15}$$

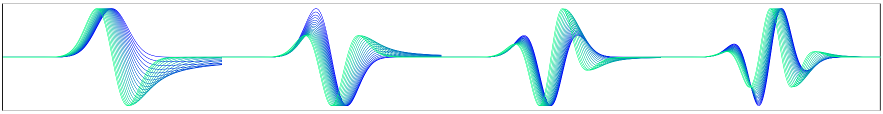



Figure 1: Fractional Order Derivatives of Gaussians, implemented using the Caputo-Fabrizio closed-form. Results are the same as those obtained by authors in [1]

.

where ${}^{CF}D_x^{n+\nu}G(x)$ denotes the $n+\nu$ order Caputo-Fabrozio derivative of $G(x)$ with respect to $x$. $n$ is the integer part of the order and $\nu$ is the fractional part. The first term on the right hand side is the integer-order part, which is a straightforward integer-order derivative and can be computed using the Hermite polynomials. The second term is the fractional-order part of the derivative which can be expressed using the Caputo-Fabrizio derivative form [6]. The CF derivative of an arbitrary function $f(x)$ can be expressed as:

$$ {}^{CF}D_x^{\nu} f(x) = \frac{1}{1-\nu}\int_0^x \exp\left(\frac{\nu}{1-\nu}(x-\tau)\right)\partial_x f(x)\Big|_{x=\tau} d\tau \tag{16} $$

Therefore, the CF derivative of a Gaussian can be expressed as:

$$ \begin{aligned} {}^{CF}D_x^{\nu} G(x;\mu,\sigma) &= \frac{1}{1-\nu}\int_0^x (\mu-\tau)\exp\left(-\frac{\nu}{1-\nu}(x-\tau)\right)\left[\frac{\mu-\tau}{\sigma^2}G(\tau;\mu,\sigma)\right] d\tau \\ &= \frac{1}{(1-\nu)\sqrt{(}2\pi)\sigma^3}\int_0^x (\mu-\tau)\exp(-\frac{1}{2\sigma^2}\left(\tau-\left(\frac{\nu}{1-\nu}\sigma^2+\mu\right)\right)^2 \\ &\quad -\frac{\nu}{1-\nu}\left[x-\mu-\frac{\sigma^2}{2}\left(\frac{\nu}{1-\nu}\right)\right])d\tau \end{aligned} \tag{17} $$

To make the equation simpler to read, we can separate out the integral into a new term that we call $\zeta_{\mu,\sigma,\nu}(x)$

$$ {}^{CF}D_x^{\nu} G(x;\mu,\sigma) = \frac{1}{1-\nu}\cdot\frac{1}{\sqrt{2\pi}\sigma^3}\exp\left(-\frac{1}{1-\nu}\left(x-\mu-\frac{\sigma^2}{2}\cdot\frac{1}{1-\nu}\right)\right)\zeta_{\mu,\sigma,\nu}(x) \tag{18} $$

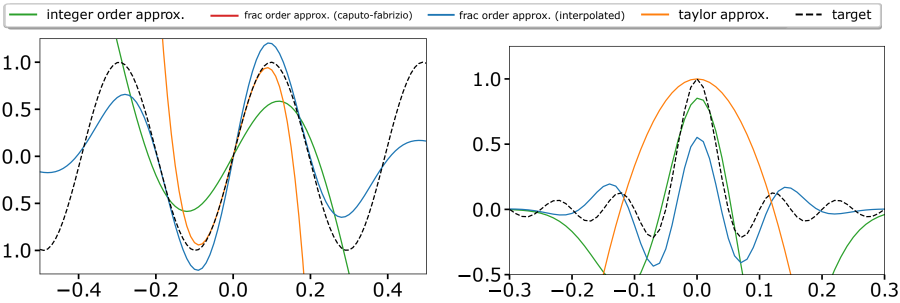
$$ \zeta_{\mu,\sigma,\nu}(x) = \int_0^x (\mu-x)\exp\left(-\frac{\left(\tau-\mu+\frac{1}{1-\nu}\sigma^2\right)^2}{2\sigma^2}\right)d\tau \tag{19} $$

To implement the fractional derivative in closed-form, we need to approximate the integral $\zeta_{\mu,\sigma,\nu}(x)$ using cumulative trapezoid or adaptive quadrature integral approximation methods [2]. The fractional derivatives of Gaussians computed using the CF form can be seen in figure Fig. 1 which are the same as results obtained by the authors. We can observe that the fractional-order smoothly transition between the integer orders which is a crucial requirement.

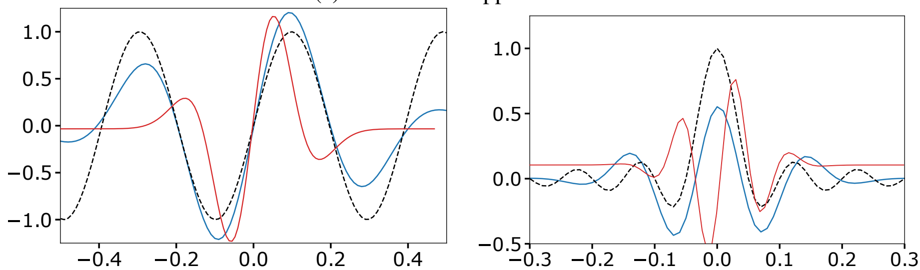
# 2 Additional experiments

## 2.1 Approximation of 1D functions

**Experimental Setup.** Before we embed our method into a CNN, we would like to test our proposed scaled fractional Gaussian derivative in a simpler 1D setting. To do this, we create




(a) 1D Function approximation



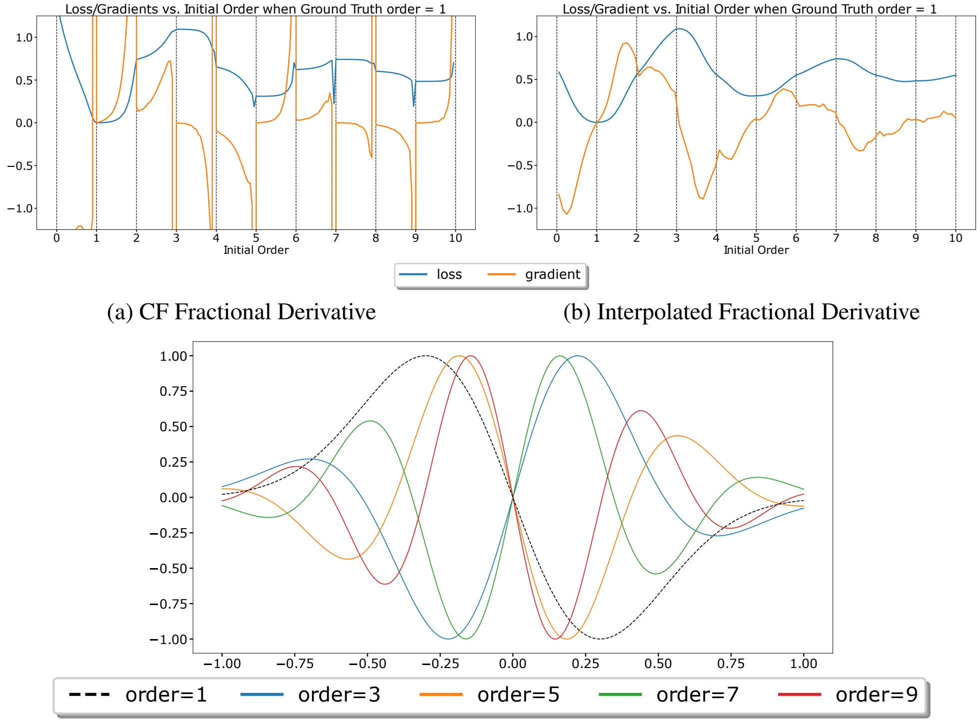
(b) Caputo-Fabrizio vs. Interpolated Fractional Derivative

Figure 2: **1D function approximation.** (a) Approximating two target functions (dashed lines) when using: a Taylor approximation, a linear combination of integral Gaussian derivatives, and a scaled fractional Gaussian derivative (computed by interpolating between integer orders). Our fractional Gaussian derivatives can accurately approximate the function. (b) Comparison between the approximations by scaled fractional Gaussian derivative computed via the Caputo-Fabrizio formulation and the interpolated method shown in (a).

a test scenario where we have 1D target functions which we would like to approximate with a fractional Gaussian basis of a single learned order (our approach) and compare this with the standard Taylor approximation of a function and previously proposed linear combination of integer order Gaussian derivatives as used in [5]. While the 'integer order' method learns the optimal coefficients for the fixed basis set required to fit the function, the 'fractional order' method learns a single order. Both methods learn via gradient descent to optimize a squared error loss. The Taylor and the 'integer order' method approximate the function with 3 terms.

We hypothesize that our approach is better at approximating the functions than other approaches because the other methods rely on a large number of terms (in the case of Taylor approximation) or orders (in the case of the linear combination of Gaussian derivatives), while our method is can learn the correct order via gradient descent. This is more true for functions with a higher frequency which can still be approximated with a single learned order rather than relying on a large number of fixed terms. We experiment with two different variations of our approach. The first by making the fractional-order derivatives using the Caputo-Fabrizio form and the next by the interpolating between integer orders.

**Accuracy of Approximation** In Fig. 2. (a) can be seen how the interpolated variant of our method performs against the integer-order and Taylor approximation. We show the results here for the sinusoid and the sinc function where our method can better approximate the functions than the other methods. While the other methods are accurate only towards the centre (which is expected), our method also performs relatively well a few standard devia-




(a) CF Fractional Derivative

(b) Interpolated Fractional Derivative

(c) Aligning of peaks of odd-ordered Gaussian Basis to the Ground-Truth order (dashed line)

Figure 3: **Gradients analysis.** Loss and gradient landscape of CF fractional derivative: (a) and 'interpolated fractional derivative' method; (b) methods for different initialization of order when approximating a Gaussian basis of order = 1 (ground truth order). The reason for the local minima in odd orders in the loss landscape is that they are aligned with the peaks and troughs of the odd order Gaussian derivative as shown in (c).

tions away. This result is promising as we expect this to translate into better results than SRF networks in 2D where the filters made of our Fractional Gaussian basis will try to approximate local image functions.

**Caputo-Fabrizio (CF) vs interpolated fractional derivatives** In Fig. 2. (b), we show some comparisons between the approximation of the CF and the Interpolated Fractional Derivative. We consistently noticed that the performance of the CF variant is lower than the interpolated one. Additionally, we noticed that the CF Derivative is highly sensitive to the initial conditions and would sometimes never update the order. Sometimes, we observe a numerical underflow of the gradients, and so we needed to clip the gradients. We investigate why such errors happen in the next section.

## 2.2 Gradients of fractional derivative methods

**Experimental Setup.** In this section, we analyse the gradients and loss landscape of the CF-Derivative and 'interpolated derivative' methods. This is in an effort to understand why the CF Derivative method is so sensitive to the initial order and the final approximation is far below the 'interpolated derivative' method. This is even more surprising considering that both derivatives smoothly transition between orders of Gaussian basis functions.

We initialize both methods at different fractional orders ranging from 0 to 10. At each of these initial conditions for both methods, we record the value of the loss function and the gradient of the loss with respect to the order. Importantly, we do not perform any gradient descent steps. The function that we task each method with approximating is simply a Gaussian derivative of arbitrary order which we call the Ground Truth order. In doing this, we would like to eliminate all external factors that could influence the difference in performance between the two methods. We also do not learn the coefficients of the basis for the 'integer order' model and make sure that the target Gaussian is scaled appropriately. We plot the gradient and loss values for each initial order condition. We expect that when the initial order matches the ground-truth order, the loss goes to zero. Additionally, to the left and right of the optimal order, the loss should be convex. We would hope to see smooth gradients that do not grow suddenly to infinity or diminish to zero.

**Analysis of Gradient and Loss Landscape.** In figure Fig. 3, (a) and (b) we see the plots of the loss and gradients with respect to different initial orders. The gradients and loss landscape of the 'interpolated order' method are as described and expected for a well-behaved method, while the gradients and loss functions of the CF fractional derivative method are not so well behaved. We notice exploding gradients while the initial order increases and approaches an integer order. This lends an explanation for the numerical underflow explained in the previous section. The loss function is not exactly smooth but has several peaks and sharp drops which could explain why the method was so sensitive to initial conditions. Another interesting phenomenon is the presence of the sharp drops close to odd integer locations which can be explained by the fact that the peaks of odd-ordered Gaussian derivatives align to reduce the loss. This can be seen illustrated in Fig. 3.(c).

# References


[1] Jorge M. Cruz–Duarte, Juan Rosales–Garcia, C. Rodrigo Correa–Cely, Arturo Garcia–Perez, and Juan Gabriel Avina–Cervantes. A closed form expression for the gaussian–based caputo–fabrizio fractional derivative for signal processing applications. *Communications in Nonlinear Science and Numerical Simulation*, 61:138–148, 2018. ISSN 1007-5704. doi: https://doi.org/10.1016/j.cnsns.2018.01.020.

[2] Walter Gander and Walter Gautschi. Adaptive quadrature—revisited. 40(1):84–101. ISSN 1572-9125. doi: 10.1023/A:1022318402393. URL https://doi.org/10.1023/A:1022318402393.

[3] J.F. Gómez-Aguilar, Teodoro Cordova, Jesús Escalante-Martínez, C.M. Calderon-Ramon, and R. Escobar Jiménez. Electrical circuits described by a fractional derivative with regular kernel. *Revista Mexicana de Fisica*, 62:144–154, 03 2016.

[4] Jordan Hristov. Transient heat diffusion with a non-singular fading memory: From the cattaneo constitutive equation with jeffrey's kernel to the caputo-fabrizio time-fractional derivative. *Thermal Science*, 20:19–19, 07 2016. doi: 10.2298/TSCI160112019H.

[5] Jorn-Henrik Jacobsen, Jan van Gemert, Zhongyu Lou, and Arnold W. M. Smeulders. Structured receptive fields in cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[6] R. Khalil, M. Al Horani, A. Yousef, and M. Sababheh. A new definition of fractional derivative. 264:65–70. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2014.01.002. URL https://www.sciencedirect.com/science/article/pii/S0377042714000065.