Fourier-Taylor Features for Median Dimensional Neural Network Functions Mr. Anomitra Sarkar, Ms. Lavanya Vajpayee

*School Of Computer Science Engineering,*

*VIT Chennai, India*

***Abstract***— In this paper, we propose a novel approach to representing neural network functions using the Fourier and Taylor series expansion. By leveraging the properties of the Fourier transform and the Taylor series expansion, we can capture the underlying structure of moderately complex problems and decompose to a functional output in a more efficient and interpretable manner. Our proposed method, termed Fourier-Taylor features, offers several advantages over traditional methods of representing neural network functions. It provides a framework for generating new neural network architectures that are optimized for specific tasks or applications. We demonstrate the effectiveness of our approach through experimental results on several benchmark datasets and compare the performance of Fourier-Taylor features with traditional methods of representing neural network functions. Our results show that Fourier-Taylor features offer improved accuracy and efficiency in representing and analyzing neural network functions and have the potential to significantly impact the field of artificial intelligence and machine learning in especially scientific studies.

***Keywords:*** Fourier Features, Taylor Features, Environmental Radiation Spectra, Convolution Neural Network, Image Segmentation and recognition

Furthermore, by incorporating Fourier series' frequency-domain analysis and Taylor series' polynomial representations, neural networks can better model dynamic systems, filter noise, and reconstruct signals.

The question does arise why Fourier series out of all? Since Fourier series shows much more promises in field of image and signal analysis. Also discussed in the latter sections we can see the obvious choice is Fourier and since its similarity to Taylor series in the Complex domain

This study investigates the theoretical foundations of Taylor-Fourier series-based image prediction, demonstrating its superiority over traditional compression methods and polynomial-based image masks. Whilst the objective is not only to suggest a replacement for the existing models for low-level networks since for high-level network the difference is insignificant, but it is to suggest a unified model/methodology for scientific and mathematical modelling and sampling. As this model produces a for low parameterization till a median level network with equivalent or even lower parameterization. The proposed framework promises to revolutionize lightweight image prediction, facilitating efficient, high-quality image analysis for diverse applications.

1. **INTRODUCTION**

In Mathematics approximations is a tool used since our very history itself. A mathematical approximation, such as Fourier or Taylor series, is a way to represent a function using a sum of simpler functions, called terms. The idea behind these approximations is to break down a complex function into smaller parts that are easier to work with, and then use those parts to estimate the value of the original function at a given point. There are different approaches too for approximation, techniques like Binomial, Lagrange and Neumann approximation which also provide a better approximative answer but in regards to concepts like simplicity and use-cases, no approximation bodes similarity to the Taylor Series and Fourier Series.

The rise of digital technology has led to an exponential surge in image data, warranting efficient compression techniques to facilitate seamless storage and transmission. Conventional methods, such as JPEG compression, also rely on discrete cosine transformations, which often compromise image quality, but nevertheless is the most popular engine and compression methodology to compress an equation based of parameters and equations. Also recently, polynomial-based image masks have emerged as a promising alternative, but their non-continuous edges hinder further analytical processing, particularly in differential and integral studies. To address this limitation, this research explores the union of Taylor and Fourier series in lightweight image prediction, offering a novel, optimal approach.

By leveraging the approximation capabilities of Taylor series and the frequency-domain analysis of Fourier series, our methodology generates smooth, continuous polynomials, eliminating sharp edges and preserving image integrity. A Fourier transform allows us to dismantle an image to its frequency domain just like it is done in Discreate Cosine Transform or the DCT. In this domain, we can remove insignificant frequencies, which basically means stripping data from the image. We can then reconstruct the image with less frequencies to obtain a "compressed" image.

This enables accurate differential and integral analysis, by boosting function approximation, signal processing, and pattern recognition capabilities.

**II. Related Works**

Recent advancements in neural networks have led to their widespread adoption in representing various visual signals. Notably, researchers have explored efficient neural network architectures for image analysis and scientific modelling. Several studies have investigated the use of Fourier feature mappings to enhance neural network performance. For instance, Tanick et al. (2020), see Ref[1] proposed employing Fourier feature mappings to transform input sequences before passing them to neural network layers, enabling the learning of higher frequency content. Their findings demonstrated that random Fourier feature mappings with appropriately chosen scales yield improved performance across various low-dimensional tasks.

Our work builds upon these insights, aiming to design an optimized Neural Network architecture that achieves substantial parameter reduction compared to conventional architecture while maintaining comparable performance for median-level image recognition tasks.

It’s a experimented concludion that Fourier features are better for Machine learning and pattern learning, especially they work pretty well when to predict and random image even a non DCT based well-defined function. But they are susceptible to overfitting of data since the amplitude fluctuations are typically the one that tends to fit in the data noise and seem to overfit. This happens for low frequency and small dataset however this is not a problem with Taylor features, which fit well without any regards to frequency and dataset size.

To leverage the benefits of Fourier features while mitigating their trade-offs, we combine them with Taylor features. This hybrid approach ensures desired output without sacrificing performance. Our optimization algorithm's non-local search capability facilitates this combination. But why does Fourier and Taylor series work in a better efficiency?

The efficiency of Fourier and Taylor series can be attributed to their recognition of feature functional attributes as mixed tokens from the feature layer, which are then improved based on metrics like accuracy and precision or simply a loss function. This concept is supported by FNet, see Ref[], which utilizes mixing tokens via Fourier transforms to speed up transformer encoder architecture.

In the FNet architecture, attention sublayers were replaced by simple linear token that can mimic the behavior of non-linear features. FNet achieved improved performance with reduced parameters. Fourier transforms exhibit a small learning and error threshold, making them suitable for mathematical modeling. Using them is a clearly effective since just like using activation functions ReLU that convert the incoming input to a non-continuous yet linear function, (or in some case like a Leaky ReLU, see Ref[], that reduce the negative counterpart by some constant instead of nullifying it to complete zero), to exhibit small non-linear/polynomial behavior in the entire network. In the similar manner the inputs here are fed to the feature layer to have a mix of both linear and non-linear characteristics.

This can have a more outpacing effect when using mathematical models, which is reflected as a test against benchmark model, see Ref. [Mathematical Modelling of Environmental Radiation Spectra using Convolutional Neural Network]. Mathematical modelling of the this CNN includes CNN and custome filtering options against various weights and hyper tuned parameters. This contributes to the advancement of computational methods in environmental science and offers potential for further research and application in environmental spectral analysis using CNNs. This model use Savitzky-Golay Filtering and Pyramid Pooling layers for its analysis of environmental spectra that offers potential for further research and application in the scientific and mathematical domain.

Sampling and processing data in Scientific and mathematical domains regarding practical applications is relatively harder since the data is limited to certain edge cases making the data sparse and doesn’t bestow a lot of data onto the Deep Learning model. Therefore making the need of Neural networks architecture less Data less and highly accurate. Therefore the models made are heavily dense to only provide metrics that are well in the boundages of considerable. But as the evolution of more scientifics sensing projects, the models require less dense field architecture making it a tricker options.

Whilst this model does perform good, the architecture is susceptible to further change to improve accuracy compared to previous model, this makes this result inefficient in the broader view of providing a model that supports minimal architecture change without drastically booming the baseline trainable parameters. On the other hand, our model leverages prediction by not data mapping as opposed to what convential architecture do. Rather we predict the fourier and taylor combination polynomial function mapping for prediction.

**III. Methodology**

As described in the previous section our method, our methodology includes functional prediction rather than data prediction. Since most of the data around have an underlying function basis which is even visible when mapped correctly. This is the essence of traditional machine learning algorithms like regression training, see Ref[], where any bivariate relationship could be shown as a linear dependence and can be used to create a supervised trained numerical prediction. Yet the current the Deep learning revolves around input and output feeds and tweaking the internal parameters, weights and biases to get closer to a certain output. So as to not functionally map it but to predict a blind data set and produce results based off on previous variables and feeds provided.

The entirety of neural networks can be clubbed under as data approximators, or better called Universal Approximators. Now they use a better optimizer to give answers to complicated tasks. A higher order example would be how a Generative AI works, which would be taking in the prompt and predicts the next string of token to create a meaningful sentences, the rest of architecture is to enhance the performance by tokenization, vectorization, etc, see Ref[].

Our methodology consists of using a combination of discrete Fourier and Taylor feature matrix for a certain order, the matrix allows values within the entire Real domain. This matrix is a Fourier-Taylor combinatorics matrix that is a description of the equation's constants. These constants include both magnitude and phase difference of the entire term.

Instead of using a CNN to get the feature extraction or get the image prediction faster compared to a general feed forward neural network we use a normal neural network that uses a FT map and can perform at a same level as a CNN model.

So, why not to use a CNN model?

CNN Model gives more accuracy compared to normal neural network for task such as object detection and image segmentation. It is significantly slower than traditional model for any one epoch, why so? Because of operation such as Pooling, maybe Max Pooling or Average Pooling to name so. and calculating filter values for each convolutional feature map. Whilst a normal neural network tries to learn and approximate the weights to get to the global maximum for optimizer function but for 2D task it fails to deliver the same accuracy that a CNN model does for slower speeds.

Our methodology suggests the same neural network where the variables are not independent terms to give out a certain prediction but rather use an initial combinatoric matrix that represents a Fourier-Taylor combination equation where the terms are calculated independently.

In simpler terms whilst a CNN model breaks down the entire image multiple times and processes extracting features out of various tensor. Our mechanism does this costly mechanism to break the image into its functional components and predict them separately in a similar way a dense neural network is trained.

The below equations is the representation for a Taylor series for a variable x and approximation point near by “a”.

Since we restrict the complex variable which is the real axis, we can constrict the complex variable to a unit circle in the form If we expand this equation and formalize it in hyper space plane, or in the extra added dimension, we use the expansion, , then we obtain the Fourier series,

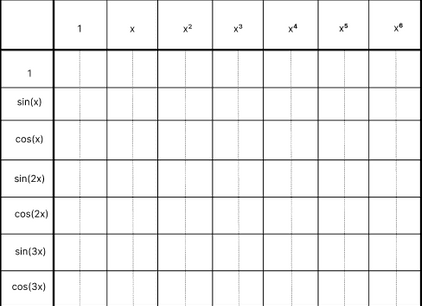
The actual use case in Fourier Taylor series boils down to addition of a hybrid matrix instead of simply using the previously tested model, see Ref[]. Since the fourier series is highly affected by dimensionality of the problem and data overfitting. Addresssing the data overfitting we can clearly say any equation that is a wave form if not for a low amplitude might just fit for very sparse data, such that it sets on a median points. Fourier series seems to be very prone to it. Therefore to reduce it we use the Taylor features which add two things.

1. Since for data points that are basically overfitting under the wrong function taylor approximation nullifies this overfitting giving a better lower amplitude and almost smooth polynomial which is majorly non-transcendental in nature. Therefore it nullifies the sparse data.
2. Since for fourier series is a taylor series mapping to higher dimensionality it takes of the extra load when the dimensionality is less and does not need a necessary input from the fourier breakdowns. This helps in managing the dimensionality by taking off the loads according to variables provided and dimensional plane needed for calculating the prediction.

This helps the methodology to not only work with speed but deliver the same modelling. We propose the algorithm to be used in image segmentation for light weight system.

This scenario compares similarly to a difference and tradeoffs a transformer to a mamba-based architecture.

Using a combination of both Fourier Series and Taylor Series helps us to contain the tradeoff of both the use cases and increase the efficiency of Fourier features from Low dimensional to Median Dimensional Problems without making frequency of distribution a problem.



**The matrix is FT-combinatorics Matrix that provides the information for an equation that contains a combination of both Fourier and Taylor Series to provide better insights.**

For the nth order the matrix thus created has dimensions as:

**(n+1)\*(n+2)\*2 matrix.**

Instead of using a convolutional network and predicting feature maps onto convolutional layers, we suggest an alternative to predict Fourier-Taylor features to predict once and compare it to the output. The two channels are used to predict the two constants in a term that effects the function as a whole:

1. One being the magnifying constant that amplifies the function/slopes
2. Second being the coefficient that translates the variable and effects the phase difference of the function.

The first row of the matrix is the kernel that is purely affected by Taylor Function.

The first column on the other hand is purely Fourier effected function.

The 2 channels are used to predict the  
two constants contained in a term that   
affects the functions as the whole:

The entire kernel prediction is based off on a mixture of both this can be adding more channels.  
Having Multiple Channel can directly affect the combinatorics matrix and equation. The dimensionality can be taken as one for any one variable and not decompose it to entire series again (since as order inside a transcendental term has different equation expansion that creates a relatively higher overhead). This creates parallel computation of different variables by not increasing time and solving the dimensionality problem to a certain extent, keeping it still unsuitable for very high dimensional problems but fairly good for moderate and median dimensional problems. As it includes both Taylor Series and Fourier series to solve dimensionality and sparse data nullification.

This methodology does have a caveat, which we do see in the latter observations section. Since every functional component for a images can create a lossy breakdown some features might get lost, though minimal it does make the prediction correct with a relatively lower accuracy than a CNN by average 10% accuracy loss.

Just like using DCT breaks down the image into 8x8 sections so as to decomposes in quantized coefficients that reduces the size of transmission, but does result in a relatively lossy decomposition, i.e, when these coefficients are joined again they either result in redundant or incoherent values correlation. In a similar way, our methodology does have a lossy decomposition and processing but it is a minor difference since the image is not made to join again.

**IV. Algorithm**

Our Algorithm includes the following steps so as to be implemented the concept and results covered in the scope of the paper

1. Creating a dense layered network model like a forward feed neural network with any optimization functions or activation functions such as ReLU or Sigmoid.
2. Adding a initial layer of Fourier-Taylor Function Feature Mapping with any no of order that is thought of. any order say "n" will produce a matrix of "n" x "n" and will have a internal channel of size 2 that contain the constant values respectively.
3. This layer should be implemented after flattening the input, for any shape of array whether be a 1D/2D/3D it should flattened out a single 1D array and pass through this filter. This is layer that communicates the feed forward and back propagation
4. The given model can be used for any output fit function accordingly.

It is as simple as saying that the model is predicting the appropriate function for the input and fine tune the equation and matrix as it gets closer to a global optimal solution.

The feature mapping implementation is provided in the GitHub link for further testing.

[Explain the working of coefficients here]

Lets take the methodology table for Fourier Table Map yet again assuming these are the coefficients and has been transformed against a arbitrary function to produce these.

A grid of squares with numbers

Description automatically generated

This network gives the following output and is responsible to give the output function, when computed in a (-15,15) for the X variate.

**V. Implementation**

Before concluding any results our testcases would be to test our implementation on the domain of general applicability and then narrowing the domain to scientific and mathematical modelling.

The general applicability domain includes test against:

1. Analysis against a function compared to neural network
2. Analysis against image dataset compared to NN vs CNN vs FTNN
3. Analysis against image segmentation masking CNN vs FTNN
4. Parametric analysis in modelling 3 Dimensional function compared to NN

For the showcasing the use case that ranges in science, we will be testing it against, Environmental Radiation Spectra CNN as provided in the referenced paper.

A graph of loss curves

Description automatically generated

For the same number of epochs the Fourier Taylor Neural Network performs better than the traditional neural network by a world of difference.

[Model Compare params 180K Parameter model for a function approximator in 1D, write it better]

[Conclude the features algorithm implementation and actual use case identify as the better model for use in scientific usecases.]

Algorithm division and variable explanation in regards to accuracy] [Add references that are related to Fnet, Fourier division, and Stochastic mechanism, SGD / ADAM optimizers, Complex domain and physics domain relation with Fourier and Taylor approximations. ] [Reference addition and proper description Stovacky-Pyramid Filtering and Hyper Environmental Sprectra analysis and General benchmark against BERT datasets.] [Quantization in regards to trade off of JPEG and here]

[Add acknowledgement if required towards SCOPE and VIT and SandyBoi]

1. *Table Captions*

Tables must be numbered using uppercase Roman numerals. Table captions must be centred and in 8 pt Regular font with Small Caps. Every word in a table caption must be capitalized except for short minor words as listed in Section III-B. Captions with table numbers must be placed before their associated tables, as shown in Table 1.

1. *Page Numbers, Headers and Footers*

Page numbers, headers and footers must not be used.

1. *Links and Bookmarks*

All hypertext links and section bookmarks will be removed from papers during the processing of papers for publication. If you need to refer to an Internet email address or URL in your paper, you must type out the address or URL fully in Regular font.

1. *References*

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES

1. S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed.,

R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

1. J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
2. S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
3. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in *Proc. ECOC’00*, 2000, paper 11.3.4, p. 109.
4. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital- to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.
5. (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
6. M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-> archive/macros/latex/contrib/supported/IEEEtran/
7. *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
8. “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
9. A. Karnik, “Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP,” M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
10. J. Padhye, V. Firoiu, and D. Towsley, “A stochastic model of TCP Reno congestion avoidance and control,” Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
11. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.