Proceedings of CONF-MPCS 2025 Symposium: Leveraging EVs and Machine Learning for Sustainable Energy Demand Management
DOI: 10.54254/2753-8818/2025.GL23412

Convolutional Neural Networks from the Perspective of Fourier Transform

Ziyuan Huang

JinQiu International High School, Jinan, China
aa43630299@outlook.com

Abstract: The learned convolution operation process of convolutional neural networks (CNNs) can be understood as a Fourier transform, or Fourier series expansion. Essentially, this means it is the weighted sum of different orthogonal vectors. The convolution kernel can be seen as understanding the dimensional orthogonal vector. This perspective is not only insightful for grasping the essence of CNNs but can also be further helpful to understand the position coding of transformers. At the same time, CNNs have the potential to be combined with signal processing techniques. This combination can improve the traditional signal transmission mode to a certain extent by providing more efficient and robust methods for signal analysis and processing. This paper will explain the basic principles of CNNs from this new perspective. It aims to help people see why CNNs can be effectively applied in the field of signal transmission, offering a fresh understanding of their capabilities and potential applications in this area.

Keywords: Fast Fourier transform, Convolutional neural networks, Long Short-Term Memory, Piecewise function

1. Introduction

Artificial intelligence is now the dominant technology, and the main technology of artificial intelligence is convolutional neural networks. Significant progress has been made in image recognition, and the application of genetic algorithms has been considered. Typically, the diagnosis begins with a fault tree that is developed based on the physical behavior of the electrical system under consideration. This is mainly due to the availability of large-scale annotated datasets and deep convolutional neural networks (CNNs) [1]. Obviously, convolution is at the heart of convolutional neural networks. People can find formulas for convolutional neural networks in most advanced math textbooks, but if people can understand these complex formulas through the concept of Fourier series, one can easily understand the concept and apply it to techniques of signaling. Traditional methods of signal transmission are slower and less secure.

To proceed, the author shall have a brief overview of the literature on feature extraction and signal classification. Running and committing from the entire notebook on an interactive session is problematic due to memory issues with the test set. Using the medium smoothing technique, the fast Fourier transform (FFT) 800.000 measuring point is used for the 8712 signals [2]. The signals are at three levels, so there are 2904 different instances of the signal. For the three signs of faith, the sum is zero. When one fails, the other continues to transmit current. It can be rectified as direct current.

DOI: 10.54254/2753-8818/2025.GL23412

In the event of a malfunction, ripples in the grinding may be observed. What is partial discharge? The typical PD is the imagine internal cavities/voids or impurities in an insulating material. When high voltage is applied on conductor, a field is also induced on the cavity. Further, when the field increases, this defect breaks down and discharges different forms of energy which result in partial discharge. This phenomenon is damaging over a long period of time. It is not event that occurs suddenly. Classical Modes of Detection Partial Discharges can be detected by measuring the emissions they give off, including ultrasonic sound, transient earth voltages (TEV and UHF energy). Is it possible to enhance the modes of Detection with better extraction of classifier characteristics? The Intel Mobile ODT Challenge 2017 aims to go beyond traditional Pre-trained CNN models for automated feature extraction [3]. There are two possible approaches, FE the signal and feed it into the NN for classification. Other NNs are used as feature extractors, and then a shallow classifier (XGBoost) is used for binary classification tasks. These include classification of long-term defects of blanket conductors based on signal characteristics, and extraction of features from time series data for classification.

2. Comparison of convolutional neural networks and fourier transform

Convolution in the time domain is equivalent to a pointwise product in the frequency domain. This property is widely used in convolution computation of accelerated CNNs (such as FFT acceleration algorithms), but its mathematical connotation goes far beyond engineering optimization. The most basic and obvious commonality between the convolution formula and the Fourier transform is reflected at the formula level [4]. This is the most common form of convolution formulas and the author can express Fourier transform by

$$F(x) = \int_{-\infty}^{\infty} f(T)g(x-T)dT\,. \tag{1}$$

2.1. Piecewise function and beyond

At first glance, these two formulas do not seem to be very correlated, so people can use mathematical methods to derive them. Obviously, there is a lack of a function $g(T)$ between these two formulas, which makes the mathematical framework of the two seem unrelated and that is exactly the most critical point. The key point connecting the two is a function that can be similar to a filter, which requires people to fix the time domain corresponding to the existing Fourier transform from a higher-dimensional perspective. Because sine and cosine curves are global, using sine and cosine as basis vectors (anchor points) to measure signals, the features found are naturally global features [5]. So the simplest way to break free from such a mathematical framework is to construct a piecewise function $g(x)F(n) = \int_{-\infty}^{\infty} f(t)g(t)\cdot e^{-i\omega t}dt$.

As the author mentioned earlier, the G function functions like a filter, and the formula that can perform this function is the gabor transform. Actually, it is only a movable window. However, it could not be used in only $F(t)$ and the $F(t)$ needs to be transformed in $F(n,s)$. The s is the site of windows and there is a unknown one k is the size of windows $g(t-s)$. So, $F(t,s) = \int_{-\infty}^{\infty} f(t)g(t-s)\cdot e^{-i\omega t}dt$. The mean effect refers to the average amount of change caused by changes in variables within a certain range. In economics and statistics, the mean effect is often used to measure the average impact of a variable on the overall outcome.

The size of k in the Gaussian distribution $g(x)$ represents the average effect. The mean effect refers to the average amount of change caused by changes in variables within a certain range. In

DOI: 10.54254/2753-8818/2025.GL23412

economics and statistics, the mean effect is often used to measure the average impact of a variable on the overall outcome.

The form of Gaussian function is $t' = \frac{t}{\sqrt{(1-v^2)/c^2}}$. According to Heisenberg's uncertainty principle, the product of the Gaussian window in the time domain and frequency domain reaches its minimum value, balancing time resolution and frequency resolution. The Fourier transform remains Gaussian: The Fourier transform of Gaussian functions remains Gaussian, making frequency domain analysis more concentrated and reducing spectral leakage. To conclude, it is found that $g(t-s) =\cdot \Pi a^{\frac{1}{2}} \cdot e^{-t-s^{\frac{1}{4a}}}$ where a is the variance $\frac{1-v^2}{c^2}$.

2.2. Embodiment of mathematical principle

The convolution operation process of CNN after learning can be understood as Fourier transform or Fourier series expansion, which is the weighted sum of different orthogonal vectors. Convolutional kernels can understand dimensional orthogonal vectors. Similarly, it can further help understand the positional encoding of transformers [6]. Each point in the frequency domain is a condensed representation of specific information across the entire time domain. The transformation domain achieves position independence, but the disadvantage is that the global perception ability is too strong, causing excessive fluctuations in the transformation domain. By adding a windowed Fourier transform, this global perception ability can be constrained. The Fourier transform after windowing is very consistent with the design of the CNN receptive field. This way, it is easy to discover the consistency of local features.

The difference in representation of the same problem in different dimensional spaces, a line in two-dimensional space only requires one point in infinite/high-dimensional space. v_1, v_2, v_n are orthogonal bases in high-dimensional space, and a point in infinite dimensional space is the weighted sum of these orthogonal bases under different weights. In this case, the physical meaning of the green curve in the transformation domain and the yellow curve in the time domain are different, although they have the same shape, one thinks they cannot represent each other. Dimensionality enhancement is not the goal, the purpose is to reduce the dimensionality of curves in two-dimensional space.

Mapping a low dimensional image to a point in high dimensional space, the focus is not on the point itself, but on how points in high-dimensional space can be equivalently represented using vectors. Low order data can be represented as a set of vectors in a high-dimensional space (a set of weights for orthogonal bases). From time domain to frequency domain, it is equivalent to making a coordinate change in Hilbert space (high-dimensional space). The orthogonal vector represented by dn can be represented by a specific vector $e^{(iwt)}$, which is the special coordinate system (coordinate axis) selected by Fourier transform in Hilbert space, or the selected special orthogonal basis. So, one can convert the infinitesimal accumulation into integral form.

Because the sine and cosine curves are global, using sine and cosine as basis vectors (anchor points) to measure signals, the features found are naturally global. Is it possible to modify this basis so that it only considers local conditions and not global conditions? That is to say, how should a base with a window range be represented? A simple method is to construct a piecewise function $g(x)$. In order to maintain the differentiability of the function representation everywhere, the function $g(x)$ is rewritten with a as the variance to determine the size of the window; S is the expectation to determine the position of the window; N remains unchanged, representing different modes/basis vectors selected. The transformation of the g-function using Gaussian distribution here is Gabor transformation, which does not take the size of the window as a variable, but rather a parameter. Because any specific window size can be determined by a complete basis (n).

DOI: 10.54254/2753-8818/2025.GL23412

The Fourier transform data communication system is a realization of frequency-division multiplexing (FDM) in which discrete Fourier transforms are computed as part of the modulation and demodulation processes. Features and positions are independent, meaning that if the positions are different, the representation should be the same in the transform domain and the features should be local. N represents different frequencies in the transform domain, and the orthogonal basis s represents the window position opened in the time domain. Only by opening the window at the corresponding waveform position can there be obvious features (weights) reflected in the transform domain.

2.3. Cross disciplinary application paradigm of CNN

In the field of computer vision, object detection and image generation are two important tasks. For object detection, Faster R-CNN is a powerful approach that combines a Region Proposal Network (RPN) with a CNN to achieve an end-to-end detection process. This integration allows for efficient and accurate detection of objects within images, achieving a mean Average Precision (mAP) of 59.1% on the challenging COCO dataset [7]. In the realm of image generation, DCGAN, or Deep Convolutional Generative Adversarial Network, has made significant strides. It utilizes transposed convolution to reconstruct high-resolution images, particularly excelling in generating realistic facial images with impressive Peak Signal-to-Noise Ratio (PSNR) values exceeding 30dB.

In the field of pathological diagnosis, the U-Net model has demonstrated remarkable performance in retinal vessel segmentation tasks, achieving a Dice coefficient of 0.97. This result is 23% higher than that of traditional methods, highlighting the significant improvement in accuracy. Meanwhile, in tumor detection, 3D CNNs, such as the V-Net, have shown excellent capabilities in detecting lung nodules. They achieve a sensitivity of 98.6%, while also reducing the false positive rate to just 0.3 per case. These advances indicate substantial progress in medical imaging analysis.

In the realm of visual language interaction, the CLIP (Contrastive Language–Image Pre-training) model has made significant strides by aligning image and text features through a technique called contrastive learning. This innovative approach enables the model to achieve remarkable performance, with a Top-1 accuracy of 76.2% in zero-shot classification tasks. Essentially, CLIP can effectively match images with their corresponding textual descriptions even when it has not been explicitly trained on those specific examples, showcasing its powerful generalization capabilities.

With the development of neural morphological computing and quantum computing, the next generation of CNNs is likely to present three major trends. First, brain-inspired spatiotemporal modeling will integrate hippocampal memory mechanisms to achieve dynamic scene understanding, such as real-time decision-making in autonomous driving. Second, multi-scale joint optimization will involve full-stack innovation, ranging from nanoscale chip design (such as integrated storage and computing architecture) to system-level deployment. Third, ethical enhancement will focus on embedding interpretable modules and fairness constraints to build a trustworthy AI system.

3. Importance of signal classification

Signal classification plays a key role in many fields such as communications, medical, industrial monitoring, etc. In the field of communication, accurate classification of modulated signals can ensure efficient and accurate transmission of information; In medical treatment, the classification of ECG and EEG signals is helpful for disease diagnosis and health monitoring; In industrial monitoring, it can find hidden dangers of equipment failure in time to ensure safe and stable production. For example, in 5G communications, fast and accurate signal classification can enable higher data transmission rates and more stable network connections. In the medical field, by classifying ECG

DOI: 10.54254/2753-8818/2025.GL23412

signals, doctors can detect heart diseases such as arrhythmias in time, and buy valuable time for patients to treat [8].

There are some limitations of traditional signal classification methods. Traditional signal classification methods are mostly based on the statistical characteristics and feature extraction of signals, which can have certain effects in simple environments, but they are unable to cope with complex and changeable real scenarios. For example, in a complex electromagnetic environment, traditional methods have weak anti-interference ability to signals, and are prone to misjudgment. For example, in areas with high-rise buildings in cities, communication signals are susceptible to interference such as reflection and scattering, and it is difficult for traditional methods to accurately classify signals, resulting in a decrease in communication quality.

There are also some advantages of CNN Long Short-Term Memory (LSTM) for signal classification. CNN has a strong ability to extract local features, and can effectively extract the local features of the signal through convolutional layer and pooling layer operation; LSTMs, on the other hand, are good at processing sequence data and have good memory capabilities for long-term dependencies on signals. The combination of the two complements each other's strengths and excels in complex signal classification tasks. Taking ECG signal classification as an example, CNN can extract the local waveform characteristics of ECG signals, and LSTM can capture the law of ECG signal changes over time, and the combination of the two can more accurately judge the type of ECG signal and improve the accuracy of diagnosis. The convolution operation of the convolutional layer is realized by the point multiplication and accumulation of the convolution kernel and the input data, and different convolution kernels can extract different local features. The common operations of the pooling layer include maximum pooling and average pooling, where the maximum value of the local area is selected for the maximum pooling, and the average value of the local area is calculated for the average pooling, so as to achieve down-sampling [9].

When processing speech signals, the convolutional kernel of the convolutional layer can be designed according to the frequency characteristics of the speech signal to extract the features of different frequency bands. The pooling layer uses down sampling to reduce data dimensions and improve processing efficiency. The input gate calculates the weight of the input signal according to the input data and the hidden state of the previous moment, and determines the degree of input of new information. The forgetting gate calculates the retention weight of the old cell state and decides how much old information to keep; The output gate calculates the output signal based on the current cell state and the hidden state, and outputs the final result. For example, in stock price forecasting, LSTM can remember long-term price trends and short-term volatility information through gating mechanisms based on historical stock price data, so as to predict future stock price trends more accurately. The model usually extracts features from the input signal through the CNN layer, takes the extracted feature sequence as the input of the LSTM, then models the sequence by the LSTM, and finally outputs the classification results through the fully connected layer and the classification layer. Taking the modulated signal detection model as an example, the input modulated signal first passes through multiple convolutional layers to extract local features and form a feature sequence [10]. These feature sequences are fed into the LSTM, which captures the temporal dependencies in the sequence. Finally, the fully connected layer is classified according to the characteristics of the LSTM output to determine the type of modulated signal.

Datasets which are used in various academic networks so that the framework overs the traditional theory of discrete signal processing to structured datasets, treating them as signals represented by graphs so that signal coefficients are indexed by graph nodes and the relationships between them are represented by weighted graph edges. The Fourier Transform of an 1D signal x of length n is the following:

DOI: 10.54254/2753-8818/2025.GL23412

$$F_j = \sum_{k=0}^{n-1} xe^{\frac{2\pi ijk}{n}}, \forall n = 0, \ldots\ldots, n-1 \quad (2)$$

Representing the signal which belongs to the complex space, it is roughly a sum of sinusoidal functions, and there is one coefficient per frequency present in the signal. The frequency takes the following values. If n is even, $f = [0,1, \ldots, n^2 - 1; -n^2, \ldots, -1]$. If n is odd, $f = [0,1, \ldots, n - 12; -n - 12, \ldots, -1]$.

Noise reduction algorithms typically involve several key steps to effectively reduce noise in a signal. First, the Fast Fourier Transform (FFT) is applied to the signal, converting it from the time domain to the frequency domain. Next, the algorithm calculates the frequency associated with each component of the transformed signal. Only the coefficients that correspond to sufficiently low frequencies (in absolute value) are retained, as these are more likely to represent the underlying signal rather than noise. Finally, the Inverse Fast Fourier Transform (IFFT) is applied to convert the filtered signal back to the time domain. This process helps in removing high-frequency noise while preserving the essential features of the original signal.

4. Conclusion

In summary, the author has demonstrated that Convolution is the ongoing consequence of transient behavior. LSTM can remember long-term price trends and short-term volatility information through gating mechanisms based on historical stock price data, so as to predict future stock price trends more accurately. The model usually extracts features from the input signal through the CNN layer, takes the extracted feature sequence as the input of the LSTM, then models the sequence by the LSTM, and finally outputs the classification results through the fully connected layer and the classification layer. Taking the modulated signal detection model as an example, the input modulated signal first passes through multiple convolutional layers to extract local features and form a feature sequence. These feature sequences are fed into the LSTM, which captures the temporal dependencies in the sequence. Finally, the fully connected layer is classified according to the characteristics of the LSTM output to determine the type of modulated signal. The use of CNN in signal transmission will be valuable in various fields in the future, especially in improving the efficiency of machines and the ability to capture and filter big data. LSTM units' ability of finding long relation from its input sequences as well as extracting local and dense features through convolution operations.

References

[1] *LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.*

[2] *He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). IEEE.*

[3] *Dosovitskiy, A., et al (2021). An image is worth 16x16 words: Transformers for image recognition at scale, arXiv:2010.11929.*

[4] *Caseiro, R., et al. (2012). Semi-intrinsic mean shift on Riemannian manifolds. In A. Fitzgibbon et al. (Eds.), Computer Vision – ECCV 2012 (Vol. 7572, pp. 342–355). Springer Berlin Heidelberg.*

[5] *Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging, 35(5), 1285–1298.*

[6] *Sandryhaila, A., & Moura, J. M. F. (2013). Discrete signal processing on graphs: Graph Fourier transform. In IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 6167–6170). IEEE.*

[7] *Filippetti, F., Franceschini, G., Tassoni, C., & Tavner, P. J. (2006). Recent developments of induction motor drives fault diagnosis using AI techniques. IEEE Transactions on Energy Conversion, 21(5), 994–1004.*

[8] *Chiu, J. P. C., & Nichols, E. (2015). Named entity recognition with bidirectional LSTM-CNNs. Transactions of the Association for Computational Linguistics, 3, 357–370.*

DOI: 10.54254/2753-8818/2025.GL23412

[9] Wang, L., Koch, D. D., Mahmoud, A. M., & Wang, J. (2011). Total corneal power estimation: Ray tracing method versus Gaussian optics formula. Investigative Ophthalmology & Visual Science, 52(3), 1716.

[10] Tranter, W. H., Taylor, D. P., Ziemer, R. E., & Bostian, C. W. (2007). Data transmission by frequency division multiplexing using the discrete Fourier transform. In Principles of Communication: Systems, Modulation, and Noise (6th ed., pp. 875–887). Wiley-IEEE Press.