Deep Learning Signal Reconstruction from Zero Crossings

Steven Queen[*]
Department of Computer Science
University of Maryland
College Park, MD 20742
squeen0@umd.edu

Abstract

We provide a neural network architecture for reconstructing a neural network architecture for reconstructing a noisy aperiodic signal from its zero-crossings. Logan's Theorem provides a subclass of band-limited signals that are uniquely represented by their time domain zero-crossings As solely a proof of existence it doesn't provide a method reconstruction from just the zero-crossings. Multiple methods for reconstruction in the periodic case have suggested, however the aperiodic case offers further challenges. Our reconstruction provides insight into the information deep learning systems are able to gain and learn from when provided a signals zero-crossings.

1 Introduction

1.1 Logan's Theorem

Zeros provide a large source of information about a signal's function. In the simple case of a polynomial the roots or zeros allow complete reconstruction within a constant amplitude. Logan's Theorem applies this idea to signals, with the goal being to reconstruct with information about the time of the signal's zero crossings. Logan's Theorem provides a proof of uniqueness for a subclass of band-limited signals which can be uniquely represented from their zero-crossings [Logan Jr, 1977]. Ideally this could provide a sampling criterion that can be viewed as a parallel to the Nyquist Theorem. The Nyquist Sampling Theorem provides a criterion for the sampling rate along the time axis for which a subclass of band-limited signals can be perfectly reconstructed. Logan's Theorem can be viewed as providing a criterion for a sampling rate along the y-axis of the signals amplitude for which a subclass of band-limited signals can be perfectly reconstructed. However, reconstruction currently no closed form solution with approximate reconstruction methods suffering from a lack of stability and application to aperiodic signals [Logan Jr, 1977].

1.2 Approaches

We attempt to construct a neural network architecture that addresses these reconstruction issues, improving on current non-linear optimization approaches [Slaney et al., 1994]. We first approach the issue of reconstruction instability as a preprocessing step, using a stacked denoising autoencoder to find a more robust representation [Slaney et al., 1994]. For the main task of reconstruction we tried various architectures. The first approach was with a Transformer due to its success in analyzing sequential data and ability to handle variable length input [Vaswani et al., 2017]. Another model made us of the NBeats architecture that aimed to provide time series prediction with interpretable design

[*]Undergraduate at University of Maryland in Computer Science and Mathematics.

Preprint. Under review.

with projection and residual prediction stacks [Oreshkin et al., 2019]. Finally a custom network learning a mask from a mixture of feedforward neural networks projecting onto a learned or Fourier series basis, as inspired by the NBeats Architecture [Oreshkin et al., 2019].

1.3 Goals

This reconstruction method is designed to apply to both the periodic case and the aperiodic signal case. This work makes steps toward a functioning neural network for reconstruction as compared to standard methods such as optimization via Projection onto Convex Sets optimization [Slaney et al., 1994]. The results don't yet provide an improvement of the Projection onto Convex Sets method, but do provide insight into effective improvements for design of a successful neural network. The aim of accomplishing this goal is to provide insight into the implications of Logan's Theorem on a wide variety of tasks. Accurate original signal reconstruction could extend the theoretical implications Logan's Theorem to a variety of deep learning tasks. This would shows that neural network systems which are inputted or learn zero-crossing information have the capacity to learn their representations from the full information of the signal. Examples of this include edge detection information in computer vision problems and neural auditory models using half wave rectified band-limited signals [Slaney et al., 1994].

2 Zero Crossing Signal Reconstruction

2.1 Logan's Theorem

Logan's Theorem aims to define a set of functions Z, where for all functions $h(t) \in Z$ we get that $h(t)$ can be reconstructed within a constant multiple from its sign function. That is:

$$Z\, s.t. \forall h_1, h_2 \in Z;\ sign(h_1(t)) = sign(h_2(t)) \implies h_1(t) = c * h_2(t),\ c \in \mathbb{C} \tag{1}$$

The first constraint set on Z is that it contains only octave bandpass limited signals [Logan Jr, 1977]. For the Fourier transform of a signal $\mathcal{F}(h(t)) = \hat{h}(t)$, a bandpass limited signal is one such that $\hat{h}(t)$ only has compact support in the frequency domain for a region r, where $a < |r| < b$, $0 < a < b$. This means that the signal has all high and low frequencies filtered out. For the signal to be octave bandpass limited a, b must be such that $0 < a < b \leq 2 * a$.

From Logan Jr [1977] we get the concept of the free zeros of a signal. Conceptually free zeros are the zeros of $h(t)$ that may be removed without destroying the bandpass limited property of $h(t)$. As it is removing any zero of $h(t)$ does not remove the highpass property of $h(t)$, but may destroy the lowpass property. Free zeros may be found as the common zeros of $h(t)$ with its Hilbert transform $\mathcal{H}(h(t)) = \hat{h}(t)$. The set $h(t)$ must have no free zeros, other than real simple zeros, for Logan's Theorem to apply. Simple zeros are those that only have multiplicity of one.

These two constraints – $\forall h(t) \in Z$ being octave bandpass limited and only having real simple free zeros – are sufficient to guarantee the reconstruction of $h(t) \in Z$ from its sign function, within a constant multiple, comprising Logan's Theorem [Logan Jr, 1977].

Let $B_{\inf}(\lambda)$ denote the set of functions $g(t) \in L_p$, $t \in \mathbb{R}$, with $g(\tau)$ for $\tau \in \mathbb{C}$ extended as exponential type λ. Additionally, let a, b still be the upper and lower bounds of the octave bandpass region. Then for

$$h(t) = real[f(t) * \exp(\frac{i * (a+b) * t}{2}),\ f(t) \in B_{\inf}(\frac{b-a}{2}) \tag{2}$$

its guaranteed that $h(t) \in Z$. This provides a convenient form to generate $h(t)$ with only real simple zeros, without checking against its Hilbert Transform, by randomly generating the bounded complex function $f(\tau)$

2.2 Reconstruction Methods

The nature of Logan's Theorem's proof of uniqueness makes direct reconstruction unstable to noise even just in the measurement of zero-crossings. Missing zero-crossings or noise in the frequency domain provide a larger challenge. Lets now consider $k(t) = h(t) + err(t)$, where $err(t)$ is an error function possibly independent of t. For $err(t) = 0$ except around one zero-crossing, it would no longer be bandpass, $k(t) \notin Z$, due to all zero-crossings not being free zeros. The orthogonal projection onto the set of bandpass functions $\hat{k(t)} = proj_Z(k(t))$, giving $\hat{k(t)} \in Z$, would give the numerically closest approximation of the zero-crossings of $k(t) \in Z$. However, regardless of the accuracy of the projection we get no theoretical guarantee that

$$\lim_{err(t)\to 0} sign[\hat{k}(t)] - sign[h(t)] = 0 \implies \lim_{err(t)\to 0} \hat{k}(t) - h(t) = 0$$

$$\text{where } \hat{k}(t) \in Z \; s.t. \; \hat{k}(t) = argmin_{\hat{k}(t)} ||k(t) - \hat{k}(t)||$$

Some reconstruction methods aiming to minimize the distance of the approximation have shown empirical success [Slaney et al., 1994, Curtis et al., 1985]. Whereas approaches such as Watanabe et al. [1995] aim to fix this via a robust projection.

One approach is to optimize the predicted signal via projections onto convex sets. The algorithm iterates between projecting onto the dual constraints of the zero crossings function in the time domain and the bandpass requirement in the frequency domain. This has been shown to be empirically effective in Slaney et al. [1994] where it is used to reconstruct half wave signals to the original bandpass filtered auditory signals. Another method creates a matrix representing a linear combination of Fourier series. The goal is to force the zeros, or eigenvalues, of the Fourier series to be equal to the given zero-crossings. In the case of noisy measurements the smallest singular value decomposition value corresponds to best approximation of Fourier coefficients. [Logan Jr, 1977].

2.3 Aperiodic Issues

Applying a reconstruction algorithm to the aperiodic case is particularly challenging. In the aperiodic case there would occur an infinite number of zeros. Any reconstruction would have to choose a finite time over which to measure zero-crossings and to reconstruct as if the measured segment was periodic sample. From the choice of segment length taken from full signal this can be viewed as taking discrete samples, or a comb, in the frequency domain, where the aperiodic signal is a continuous octave band-limited function. Any criteria for determining the length of the segment for periodic approximation would tradeoff the sampling rate of the frequency domain comb and the computational cost.

Additionally the theoretical application for aperiodic signals is limited due to spectral leakage occurring in practical bandpass signals. Filtering with window functions and digital signals representations result in spectral leakage outside of signals passband region. This results in "almost bandpass" signals which, while sufficient for many applications, is not sufficient for applying the results of Logan's Theorem. While there may exist a corresponding bandpass signal almost everywhere, we can't assert that there exists a corresponding bandpass signal with exactly equal zero-crossings. In the aperiodic signal case even in the absence of noise we do not have any theoretical guarantees about the signal's reconstruction.

3 Network Architecture

3.1 Denoising Autoencoders

Stacked Denoising Autoencoders (SDA) can play a dual role as a self-supervised pre-processing step to a neural network [Vincent et al., 2010]. First, and crucially to our application, the SDA optimizes learning an representation that is robust to noise. Secondly, denoising rewards extracting meaningful features. Given that our zero-crossing information represents a minimum of information which we wish to reconstruct our signal, ideally all of the information is useful. However, with relevant additive noise, such as in the frequency domain, it can learn the important dependencies between zero-crossings. Additionally, due to the variable number of zero-crossings within a fixed time period, our input will be zero padded. So our input will carry this unnecessary "information" to filter out. To maximize the information propagated we will encourage a sparse encoding, instead of reducing the encoding dimension. This will also improve its handling of zero padded input [Vincent et al., 2010].

The SDA will be trained on periodic signals, providing close to noiseless measurements. For the first layer the input can have additive noise in the frequency dimension with Gaussian noise added to the Fourier coefficients, including spectral leakage. This layer aims to extract features capturing the long-term dependencies of the data. The second layer will be trained against additive Gaussian noise from the encoding of the first, aiming to improve robustness to measurement noise. Masking noise, where some zero crossings are failed to measure, would likely make signal reconstruction infeasible, so we only use this for the third layer where the SDA has possibly learned non-crucial information. So the third layer is trained by randomly choosing a fraction of the second layer's encodings to set to zero. At all levels use the KL-Divergence loss function to encourage sparse representation.

$$L(y_{pred}, y_{true}) = y_{true} * \log \frac{y_{true}}{y_{pred}} = y_{true} * (\log y_{true} * \log y_{pred}) \tag{3}$$

At the final output of the entire model we wish to output the prediction of the signal at a sampling rate greater than the Nyquist sampling rate, representing the signal's full information. We use this output size as the dimension of each encoding, so as to learn robust representations that maintain the signal's full information.

3.2 Time Series Reconstruction

The first approach tried was direct reconstruction from a series of timestamps, corresponding to interpolations of when the signal crossed zero. This had the advantage of requiring less information as input, than the full sign function of the signal. Additionally, it gave the network more precise approximations of the zero crossings than the pure sign function. However, it required a variable length input with zero-padding and less clearly represents the relationship between the output structure. After the model predicted an output it was multiplied by the value at time zero in the original signal before computing the loss. This is because Logan's Theorem states that the original signal can be reconstructed within a multiplicative constant, so ideally this would restore the right scaling to the signal reconstruction.

The first approach used was a Transformer for the fact that it is suited to sequential data and can easily apply variable length inputs. The transformer from Devlin et al. [2018] was used, except no positional encoding was needed as the timestamps innately represent their position. When implemented this model didn't learn any meaningful representation. Additionally, neither adding or removing depth helped and it was computationally very expensive. Without a basic building block that could learn to any degree to build off of this approach was no longer used.

Instead the NBeats architecture was used, NBeats is a smaller model and allowed more problem specific encoding into our model design [Oreshkin et al., 2019]. NBeats has a block based architecture with each block learning a forecast and backcast prediction. The backcast prediction is used along with the original input as the input of the next block for residual learning, additionally the summation of the forecasts of each block is used as output prediction. This helps to provide an interpretable input. Each block consists of a basic feedforward network going that learns how to projected the input on to the block's basis. The basis for each block can be chosen as a Fourier series, a polynomial. This allowed us to encode problem related information to the model in the form of setting the blocks basis as a Fourier series over the bandpass. Unfortunately, even with added depth in form of block's this model also failed to learn meaningful reconstructions.

3.3 Mask Reconstruction

As the networks were having trouble learning, the input was instead changed to the sign function of the original signal. This was because the sign function has a clear relationship to the final output structure, so that the model can directly learn the signal reconstruction versus focusing on decoding the inputs relation to the output series. Additionally, the target output was a mask of the original input so that the model would need to retain less relative information throughout the computation. As such the mask was all positive as forced by its final output going through a ReLU activation. Again the final output was multiplied by the value at time zero in the original signal before computing the loss.

Taking inspiration from the NBeats architecture an assemble of blocks were used with the summation of their outputs being the output mask [Oreshkin et al., 2019]. Each block again consisting of feed forward neural networks learning projections onto their respective basis were used. Unlike NBeats the input to each block was the original sign function with no residual connections.

4 Implementation

4.1 Periodic Data Generation

With periodic signal generation we can get very accurate zero-crossing measurements from the generating functions. As such we use it for our pre-training for both the periodic and aperiodic models. The signal is randomly generated as a fourier series with uniformly distributed fourier coefficients over the signals bandpass. These functions can then be checked with their Hilbert Transforms for any non-real non-simple free zeros. Frequency noise can be added to the Fourier coefficients directly. Accurate zero-crossing measurements are calculated numerically with the Brent's method for root finding [Brent, 1973]. Using this process an unlimited amount of training examples can be generated.

4.2 Aperiodic Data Generation

A stochastic process provides a distribution of random functions. The generating distribution needs to provide only the subclass of aperiodic signals which Logan's Theorem applies. Logan's Theorem can be extended to unbounded functions in the case of sample functions from gaussian processes [Logan Jr, 1977]. The sinc kernel provides compact support for a bandpass region in the frequency domain.

$$K(t,t') = sinc(2 * \delta * (t - t')) * cos(2 * \pi * \xi * t)) \text{ where } sinc(x) = \frac{1}{\pi * x} * sin(\pi * x)$$

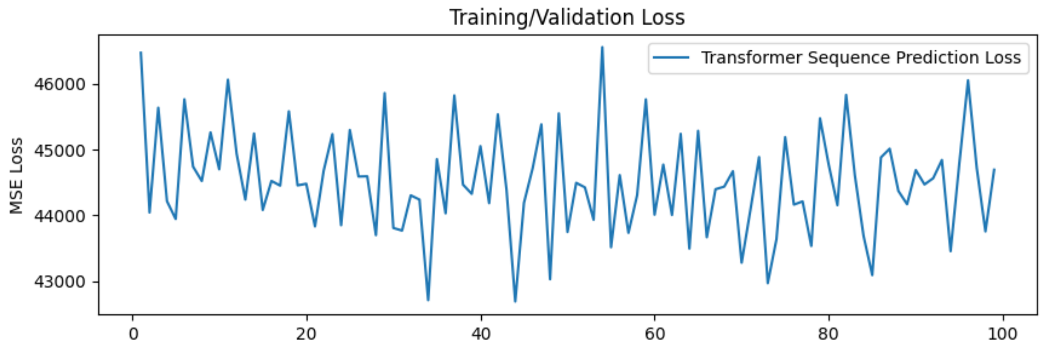
$K(t,t')$ provides a sinc kernel with compact support in the frequency domain over $[\xi - \delta, \xi + \delta]$ [Tobar, 2019]. This is used as the covariance kernel for a gaussian process to generate octave bandpass aperiodic sample functions. From section 2.1 this can be used to generate $f(t)$, for $h(t) = real[f(t) * \exp(\frac{i*(a+b)*t}{2})$, where a, b are the bounds of the chosen octave bandpass region. This guarantees that $h(t)$ is octave bandpass limited and only has real simple zeros. The zeros of $h(t)$ are identified with linear interpolation between the sample points, and the sign function can be applied directly to the output sample points. This provides an unlimited number of training examples of practical signals with measurement error and spectral leakage, from a discrete sample of the aperiodic gaussian process.

4.3 Training

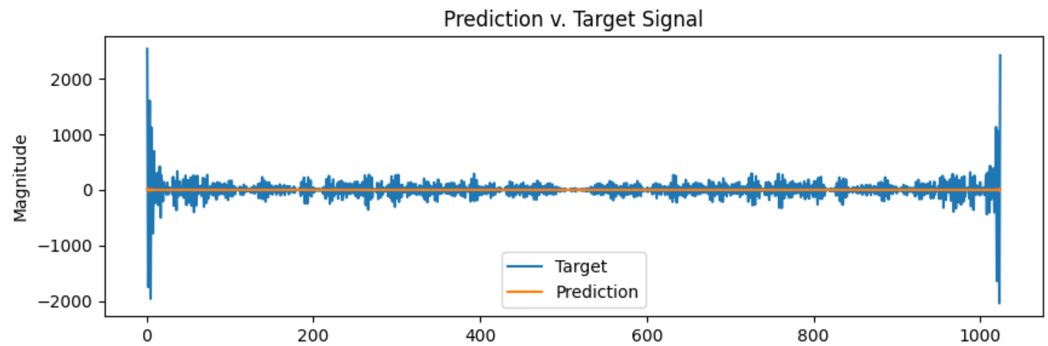
For the time series reconstruction data is inputted as a zero-padded array of ordered interpolated zero-crossing time stamps. The hyperparameters for the inputted signals (time frame, bandpass frequency, sampling rate, fundamental frequency) are constant for all training, instead of giving this information to the model. This represents a real world implementation of creating a new model for each signal generating source versus a generalized model. This has the upside of not needing to directly know the hyperparameters of the generated signal. As Logan's Theorem guarantees the ability to recover the signal within a constant multiple, the model's output is multiplied by the first value from the original signal to regain the magnitude information. Since our training data is generated artificially from our chosen generating signal distribution, the training data distribution is the same as the true data distribution. As such we can generate an infinite number of training examples, so each generated signal is used only once for training the model. Additionally, there is no requirement for a train, test, validation split as each split would be drawn from the same distribution and there is an infinite number of examples to use. All training was done in mini-batches to improve the computational speed while learning.

5 Results

The implementation of the Transformer and the block mask prediction assemble were run on the periodic training data. For both runs the optimizer use the mean squared error of the prediction



100 epochs of training of the transformer



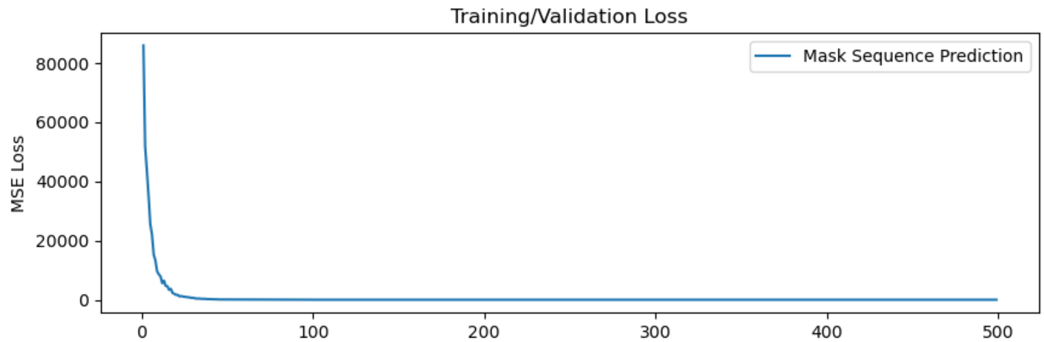
Signal predicted from Transformer after 100 epochs

against the original signal. The components of the signal generating Fourier series were between frequencies 4 and 8 hertz, with 100 regularly sampled frequencies. The signal was was evenly sampled 1024 times across one period the signal. All batch sizes were of size 20.

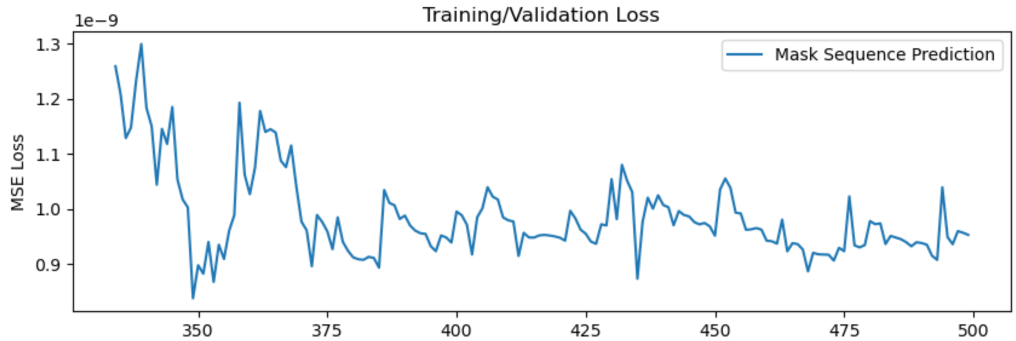
In the Transformer case the input was the interpolated zero-crossing of these discrete samples via Brent's method [Brent, 1973]. All signals had less than 800 zero-crossings across one period, so every input was zero-padded to length of 800. In an effort to improve the positional relationship of each zero-crossing to the desired sign change position in the predicted signal output the input was formatted as 800 separate 1024 vectors, each vector containing a single zero-crossing at its corresponding position in the 1024 length signal. This significantly increased training time, but was used to improve the error propagation, adding positional information to the model, as the model failed to learn for any length of learning. The transformer used had 2 layers and 2 heads. The small depth and complexity was a result of finding similar results across all variations of increased size.

For the block assemble the input was the 1024 length vector of the sign function of the original discrete signal. The assemble was made of two blocks, one projecting onto a Fourier series basis and the other onto a basis learned through the error back propagation. More than one block was found to improve results, however adding learned basis blocks yielded little improvement, while multiple Fourier basis blocks significantly hurt prediction accuracy. The feed forward network of each block had three hidden layers with width of 1000 and projected onto a basis of width 200.

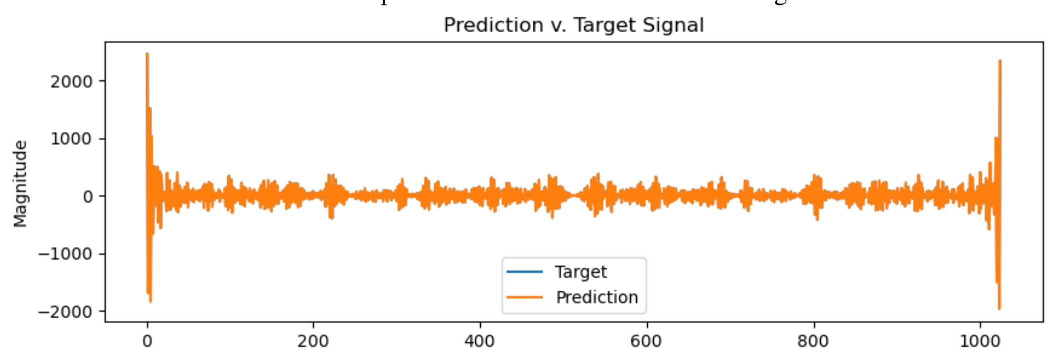
From the provided graphs on the next page, the transformer is seen to not learn at all throughout the training, increase training length did nothing to change this trend showing the same results across all lengths. Additionally, the predicted signal is seen to collapse to zero as it is unable to capture the frequency structure of the signal and unable to propagate the sign information of the signal. This problem was seen with all attempted sufficiently deep models. Conversely the block assemble shows a clearly learned signal minimizing at below a 1e-9 mean squared error. In the shown predicted signal the visual fails to show the actual signal its compared due to how tightly the prediction matches and overlaps it. Even though this prediction is done on only the periodic case, its show promises for being able to learn and apply meaningful representations from just the signal's sign information.



500 epochs of the block assemble's training



Last 170 epochs of the block assemble's training



Signal predicted from block assemble after 500 epochs

6 Conclusion

6.1 Impact

For future improvement on this signal reconstruction problem it is important to look at how to embed more problem specific information in the model. Initial attempts to improve the reconstruction accuracy through increasing depth and expressiveness of the model, failed to learn with no meaningful error signal. A more effective approach focused on a bottom up design, using simple models that could effectively learn from the data, even if it limited the types of signals the model could represent. From here incremental improvements could be made in the expressiveness of the model while still learning meaningful representations from the input. Framing the problem as learning a mask of the sign function is important as it explicitly constricts the output to one of our few known constraints and reduces the amount of information about the input signal that needs to propagate through the network. Additionally adding problem specific information such as projection onto a bandpass Fourier signal formed another effective method of embedding problem information to limit the possible output space of the model. Finally, to increase the expressiveness of these simple models using an additive assemble of these basic blocks improved the model performance while being able to effectively learn.

6.2 Limitations

First and foremost the approach needs to be fully developed, with the current work offering a road map and strengths of different approaches at tackling the challenge. At minimum the effectiveness should be comparable to the standard direct calculation approaches, before drawing insight into the impact this has the understanding of what a neural network can learn from limited signal information [Slaney et al., 1994, Watanabe et al., 1995].

For this study all of the models were trained and compared on with training samples of generated signals with fixed hyperparameters. So a single trained model works on a specific signal type, from a given signal generating source. With an improved model it would be important to evaluate the model with limited training data, looking at how at its generalizability. Additionally further evaluation would benefit from being done on relative real world examples of these signals such as auditory sound waves. However, the current approach is representative of the proposed implications of the work where neural networks are expected to learn from zero-crossing information from a problem specific subclass of the octave bandlimited class of signals. An example being the the cochlear filters of the auditory model of the inner ear learning for a specific less than octave bandpass of auditory signals [Slaney et al., 1994].

References

Richard P Brent. *Algorithms for minimization without derivatives*. Courier Corporation, 1973.

Susan Curtis, Alan Oppenheim, and Jae Lim. Signal reconstruction from fourier transform sign information. *IEEE transactions on acoustics, speech, and signal processing*, 33(3):643–657, 1985.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Benjamin F Logan Jr. Information in the zero crossings of bandpass signals. *Bell System Technical Journal*, 56(4):487–510, 1977.

Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.

Malcolm Slaney, Daniel Naar, and RE Lyon. Auditory model inversion for sound separation. In *Proceedings of ICASSP'94. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages II–77. IEEE, 1994.

Felipe Tobar. Band-limited gaussian processes: The sinc kernel. *Advances in Neural Information Processing Systems*, 32, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010.

Shinji Watanabe, Y Akimoto, Takashi Komatsu, and Takahiro Saito. A new stabilized zero-crossing representation in the wavelet transform domain and signal reconstruction. In *Proceedings., International Conference on Image Processing*, volume 1, pages 37–40. IEEE, 1995.