Super-Resolution via Learned Predictor

Sampath Kumar Dondapati [1,2,*], Omkar Nitsure[2,*], Satish Mulleti[2]
[1]*Satish Dhawan Space Center-SHAR, ISRO*, India.
[2] *Department of Electrical Engineering, Indian Institute of Technology (IIT) Bombay*, India.
Emails: mr.sampathkumar.d@gmail.com, omkarnitsure2003@gmail.com, mulleti.satish@gmail.com

***Abstract*—Frequency estimation from measurements corrupted by noise is a fundamental challenge across numerous engineering and scientific fields. Among the pivotal factors shaping the resolution capacity of any frequency estimation technique are noise levels and measurement count. Often constrained by practical limitations, the number of measurements tends to be limited. This work introduces a learning-driven approach focused on predicting forthcoming measurements based on available samples. Subsequently, we demonstrate that we can attain high-resolution frequency estimates by combining provided and predicted measurements. In particular, our findings indicate that using just one-third of the total measurements, the method achieves a performance akin to that obtained with the complete set. Unlike existing learning-based frequency estimators, our approach's output retains full interpretability. This work holds promise for developing energy-efficient systems with reduced sampling requirements, which will benefit various applications.**

***Index Terms*—high-resolution spectral estimation, learnable predictor, Transformer, LSTM, CNN.**

I. Introduction

The task of frequency estimation is to determine a set of L frequencies from a set of M measurements. This is a common problem in various applications, such as direction-of-arrival estimation, Doppler estimation in radar and sonar, ultrasound imaging, optical coherence tomography, and nuclear magnetic spectroscopy. In most applications, the frequency resolution of algorithms is a crucial factor and generally increases with the number of measurements in the presence of noise. However, the number of measurements is often limited due to cost and power consumption. Hence, algorithms capable of achieving high resolution with fewer measurements are desirable.

Theoretically, $2L$ samples are sufficient and necessary to determine L frequencies uniquely in the absence of noise. In practice, Prony's algorithm can operate with such limited samples [1]. Though the algorithm has infinite resolution in the absence of noise, it is highly unstable at the slightest noise level. Denoising algorithms could be applied to improve the noise robustness at the expense of high computations [2], [3].

The investigation centered on high-resolution spectral estimation (HRSE) has led to the exploration of multiple noise-resistant techniques (detailed in Chapter 4 of [4]). High resolution in this context denotes algorithmic precision that surpasses that of the periodogram or Fourier transform-based approach, characterized by $1/M$. Among these techniques is the multiple signal classification method (MUSIC) [5]–[7], the estimation of signal parameters using the rotational invariance technique (ESPRIT) [8], and the use of the matrix pencil method [9]. These techniques exhibit superior robustness and enhanced resolution capabilities compared to periodogram and annihilating filter methods. In [7], a precise mathematical expression is derived for frequency estimation, highlighting the inverse relationship between error and minimum frequency separation under medium to high signal-to-noise ratios (SNRs) as M tends to infinity. High resolution might necessitate a sufficiently large value of M in scenarios with non-asymptotic conditions and low SNR.

The majority of the aforementioned HRSE techniques are built upon uniform consecutive samples of sinusoidal signals. Multiple strategies have emerged for estimating frequencies from irregular samples [10]–[14]. In particular, [12]–[14] introduces an approach centered on minimization of atomic norms, which extends the concept of a norm ℓ_1 for the estimation of sparse vectors [15], [16]. Instead of using all M consecutive samples of the signal, the atomic norm-based method estimates frequencies with high probability from $\mathcal{O}(L, \log L, \log M)$ random measurements from available M measurements, provided that no two frequencies are closer than $4/M$ [12], [13]. Though the theoretical resolution ability of the method is four times lower than the periodogram, in practice, the algorithm is shown to resolve frequencies separated by $1/M$. However, these techniques have high computational complexity due to reliance on semidefinite programming.

A potential reduction in computational burden is attainable through data-driven methodologies. For instance, [17] introduced PSnet, a deep-learning-based scheme for frequency estimation. Analogous to the principles behind periodogram and root-MUSIC [7], the authors suggest an initial step of learning a pseudo-spectrum from samples, followed by frequency estimation via spectrum peak identification. Empirical observations reveal the superiority of PSnet over MUSIC and periodogram when the minimum frequency separation equals $1/M$. The important insight is that estimating frequencies from a learned frequency representation

(e.g., pseudo-spectrum) is more advantageous than training a deep network to directly learn frequencies from the samples, as attempted in [18]. An enhanced version of PSnet is proposed in [19], introducing architectural improvements in the network and incorporating a learned estimator for the frequency count L. These techniques employ triangular and Gaussian kernels to construct pseudo-spectra, although the optimal kernel selection remains an open question. Furthermore, the impact of pseudo-spectrum discretization on learning was not addressed. Diverse adaptations and extensions of these approaches are presented in [20]–[27]. In all these approaches, the choice of pseudo-spectrum or spectral representation, which is learned by the network, is contentious both in terms of its choice during training and its interpretability.

Several alternative learning strategies assume a grid-based arrangement of frequencies and employ multilabel classification to tackle the issue [28]. This "on-grid" presumption, however, imposes limitations on resolution capabilities. In a different direction, a subset of methods learns denoising techniques before applying conventional frequency estimation methods [29], [30]. These approaches necessitate broad training across various noise levels [29] or demand an extensive set of training samples [30]. In a nutshell, a non-learning-based approach requires many samples to reach a desired resolution in the presence of noise. However, the learning-based methods are data-hungry and non-interpretable, and the resolution is still not below the periodogram limit.

In this paper, we adopt a novel data-driven approach for frequency estimation from noisy measurements to achieve high resolution while maintaining interpretability. For a fixed resolution, the estimation error of any frequency estimation algorithm decreases with a higher signal-to-noise ratio (SNR) and/or an increased number of samples. Denoising can mitigate noise effects, although it possesses inherent limitations [2], [30]. Consequently, our focus shifts towards enhancing the number of samples, aiming to boost accuracy. Our proposed solution entails a learning-based predictor that augments the number of samples by extrapolating from a small set of noisy samples. Specifically, we predict $N - M$ samples from M measurements, where $N > M$. By leveraging the provided M samples alongside the predicted ones, we approach the accuracy achievable with N samples. We propose two architectures for the learnable predictor. The first is a hybrid LSTM-CNN model, whereas the second is realized by an Encoder-only Transformer [31]. Unlike methods such as [17], [19]–[25], our approach does not require the selection or design of pseudo-spectra which may not have an interpretation in terms of conventional Fourier spectra. Through simulations, we show that by using M noisy measurements, we achieve a higher resolution and lower errors for different SNRs than the HRSE approach and the method in [19], especially with the TF-based predictor.

The structure of the paper unfolds as follows: Section 2 defines the signal model and outlines the problem formulation. In Section 3, we delve into the details of the proposed predictor. Section 4 covers network design and simulation results, followed by conclusions.

II. Problem Formulation

Consider N uniform samples of a linear combination of complex exponentials given as

$$x(n) = \sum_{\ell=1}^{L} a_\ell \, \exp(\mathrm{j}2\pi f_\ell \, n) + w(n), \quad n = 1, \cdots, N. \tag{1}$$

Here, the coefficients $a_\ell \in \mathbb{R}$ represent amplitudes, and $f_\ell \in (0, 0.5]$ represent normalized frequencies. The term $w(n)$ is additive noise, which is a zero-mean Gaussian random variable with variance σ^2, with the samples being independent and identically distributed. The objective is to estimate the frequencies $\{f_\ell\}_{\ell=1}^{L}$ from the measurements.

Estimation accuracy relies on the following factors: the SNR, measured in decibels as $10\log_{10}\left(\frac{|x(n)|_2^2}{N\sigma^2}\right)$, number of measurements M, and resolution Δ_f, representing the smallest distance between any two frequencies in $x(n)$. Keeping SNR and resolution constant means that higher accuracy requires a larger number of measurements. On the other hand, for an acceptable estimation accuracy and a given SNR, the achievable resolution is around $1/N$ for HRSE methods if all the N measurements are used. When it is expensive to gather a large number of measurements and a subset of the measurements $\{x(n)\}_{n=1}^{M}$ are available, where $M < N$, the resolution decreases to $1/M$.

The question is whether a resolution of $1/N$ lower is achievable from M measurements. We show that this is possible by using a learning-based approach for frequency estimation, which will be discussed in the following sections.

III. Learnable Predictor

Utilizing HRSE and non-learnable techniques to estimate frequencies from the complete set of N measurements might yield favorable accuracy and resolution. Nonetheless, this may not hold true in cases where only M (consecutive) samples are accessible. We introduce a two-step strategy to achieve resolution closer to $1/N$ using the available M measurements. In the first stage, we predict clean samples $\{x(n)\}_{n=M+1}^{N}$ based on the provided noisy measurements $\{x(n)\}_{n=1}^{M}$. The predicted samples are represented as $\{\tilde{x}(n)\}_{n=M+1}^{N}$. Subsequently, we concatenate these estimated samples with the existing ones and employ HRSE methods to estimate the frequencies.

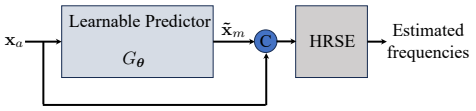


Fig. 1. Schematic of the proposed frequency estimation method: The predictor estimates the future samples from the given samples and then concatenates with the true samples by using concatenation operation C. Then HRSE methods are used to estimate frequencies from the samples.

In the initial phase, when noise is absent, it is possible to achieve perfect prediction for $x(n)$ [32]. Specifically, in the absence of noise, for any $x(n)$ as in (1), there exists a set of coefficients $\{c_k\}_{k=1}^{K}$ such that $x(n) = \sum_{k=1}^{K} c_k\, x(n-k), n = K+1, \cdots, N$, where $K \geq L$. In the presence of noise, $\{c_k\}_{k=1}^{L}$ are estimated using least-squares, aiming to minimize the error $\sum_{n=L+1}^{N} \left| x(n) - \sum_{k=1}^{L} c_k, x(n-k) \right|^2$. However, this predictive approach remains susceptible to errors in low SNR and situations with lower resolution. Furthermore, the coefficients have to be recalculated for different signals.

To construct a robust predictor applicable to a class of signals, we train deep learning models to estimate $\{x(n)\}_{n=M+1}^{N}$ based on the available samples. To elaborate, consider the vectors $\mathbf{x}_a \in \mathbb{R}^M$ and $\mathbf{x}_m \in \mathbb{R}^{N-M}$, representing the available first M measurements and the last $N-M$ samples, respectively. To be precise, let $\{\mathbf{x}_i\}_{i=1}^{I}$ represent a collection of noisy samples following the form in (1), where frequencies and amplitudes vary among examples. By splitting the measurements of these examples into the initial M measurements and the remaining set, we obtain the vectors $\{\mathbf{x}_{a,i}\}_{i=1}^{I}$ and $\{\mathbf{x}_{m,i}\}_{i=1}^{I}$. The network parameters $\boldsymbol{\theta}$ are optimized to minimize the cost function

$$\sum_{i=1}^{I} \|\mathbf{x}_{m,i} - G_{\boldsymbol{\theta}}(\mathbf{x}_{a,i})\|_2^2 \tag{2}$$

During inference, for any set of M noisy samples $\mathbf{x}_a$, the future samples are predicted as $\mathbf{x}_m = G_{\boldsymbol{\theta}}(\mathbf{x}_a)$. Then a HRSE method is applied on the concatenated vector $[\mathbf{x}_a^{\mathrm{T}}\ \mathbf{x}_m^{\mathrm{T}}]^{\mathrm{T}}$ for frequency estimation. The overall system is shown in Fig. 1.

IV. Experimental Design and Simulation Results

In this section, we first discuss network design and data generation. Then, we compare the proposed approach with the existing methods.

A. Network Architecture

1) Hybrid LSTM-CNN model: The CNN-based architecture of the learnable predictor combines convolutional and recurrent neural network as shown in Fig. 2. We employ dropout to avoid overfitting and batch-normalization is used for stable convergence.

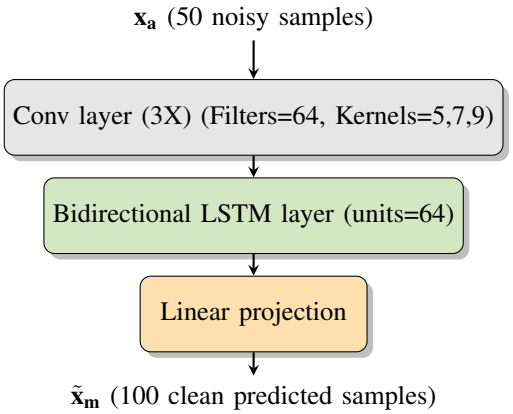


Fig. 2. Hybrid LSTM-CNN model's architecture.

2) Transformer-Encoder-based model: The model architecture is shown in Fig. 3; here, we employ batch-normalization in intermediate steps to improve convergence. Additionally, learnable positional encoding is adapted before feeding the input to the transformer block.

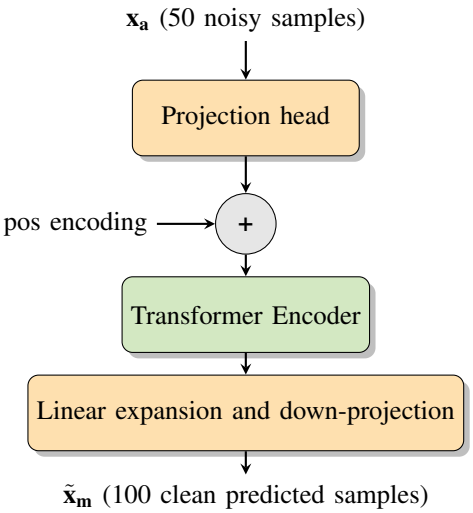


Fig. 3. Transformer model's architecture.

B. Training Data Generation

Since the task is to improve the resolution from $1/M$ to $1/N$, we considered only two frequencies ($L = 2$) with equal amplitudes during training and testing. Further, to have a resolution of at least $\Delta_f = 1/N$ and potentially improve on smaller resolutions, the samples were generated such that they had some examples with resolutions smaller than $1/N$. We generated the following six data sets to

provide a sufficient mix of frequencies for the model to generalize well.

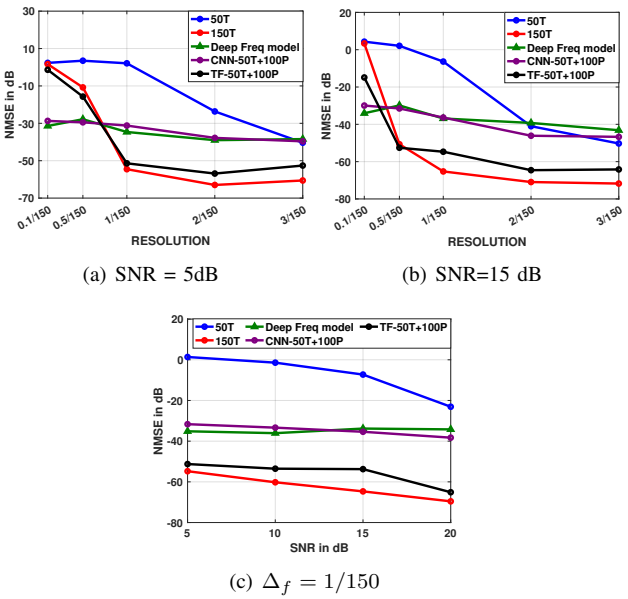
- Set-1: 20000 examples were generated with minimum resolution. For each example, f_1 was chosen uniformly at random from the interval $[0, 0.5-\Delta_f]$. Then the second frequency was set as $f_2 = f_1 + 0.5\Delta_f$.
- Set-2: In this set, 20000 examples were generated with a resolution greater than Δ_f. For each example, f_1 was chosen randomly as in the previous set. Then, we selected $f_2 = f_1 + \Delta_f + \epsilon$ where ϵ was generated from a uniform distribution $[-(f_1 + \Delta_f), (0.5 - f_1 - \Delta_f)]$. The distribution of ϵ ensures that $|f_2 - f_1| \geq \Delta_f$.
- Set-3: This set consists of 5625 examples where both frequencies were randomly generated from a two-dimensional grid with a grid size Δ_f. Since the frequencies had to be limited to the range $[0, 0.5]$, for each frequency, the grid points were given by the set $\{k\Delta_f\}_{k=1}^{75}$.
- Set-4: In this set, f_1 was chosen as in Set-1 or Set-2, and then, we set the second frequency as $f_2 = f_1 + k\Delta_f$ where k was chosen randomly from the integer set $\left[\lceil(-\frac{f_1}{\Delta_f})\rceil, \cdots, \lfloor\frac{0.5-f_1}{\Delta_f}\rfloor\right]$. A total of 20000 examples were generated in this set.
- Set-5: This set consists of 20000 examples in which both f_1 and f_2 were chosen uniformly at random from the interval $[0, 0.5]$
- Set-6: In this set, f_1 was chosen from a Gaussian distribution with a mean and standard deviation of 0.25 and restricted to the range $[0, 0.5]$. f_2 was chosen uniformly at random from the interval $[0, 0.5]$. A total of 20000 examples were generated in this set.

Set 3 and 4 were generated to help the network to separate frequencies separated by integer multiples of Δ_f, which is the key in the resolution analysis. The above six sets have 105625 possible pairs of frequencies (f_1, f_2). The samples were generated for each pair of frequencies following (1) with $a_1 = a_2 = 1$ and for $N = 150$. Each of these 105625 signal samples was corrupted with noise for a given SNR. Specifically, three independent noise instances for each example were realized, resulting in a total of 316875 examples. We also trained different networks for each SNR.

By keeping $M = 50$ (note that $N = 150$ in our setup), the network was trained for 50 epochs with a batch size of 128 in the CNN-based model and 2048 in the TF-based model. The training loss was minimized using the Adam optimizer with a learning rate of 0.001. A performance comparison of the proposed algorithm through simulations will be discussed next.

C. Simulation Results

In this section, we compared the following methods for frequency estimation by setting $N = 150$ and $M = 50$.



(a) SNR = 5dB (b) SNR=15 dB

(c) $\Delta_f = 1/150$

Fig. 4. A comparison of various methods for different resolutions at 5dB and 15dB SNR where $L = 2$: The proposed approach, labeled as TF-50T+100P, has 50 dB lower error than 50T for frequency separation equal to the resolution limit $\Delta_f = 1/150$.

- *50T*: Used $M = 50$ noisy samples with ESPRIT [8].
- *150T*: Used $N = 150$ noisy samples with ESPRIT [8].
- *CNN-50T+100P*: Used 50 true noisy (*T*) with 100 predicted (*P*) samples with ESPRIT. The predictor is realized by CNN.
- *TF-50T+100P*: Same as the previous approach, but the prediction is via the TF.
- *Deep-Freq* model [19]. The model is trained using the proposed dataset

For an objective comparison, we used normalized mean-squared error (NMSE), which was computed as

$$\text{NMSE} = \frac{\frac{1}{K}\sum_{k=1}^{K}(f_k - \tilde{f}_k)^2}{\frac{1}{K}\sum_{k=1}^{K}(f_k)^2}, \tag{3}$$

where $\tilde{f}_k$ is an estimate of the frequency f_k.

In the following simulation, we examine the model performance under varying resolution conditions for SNRs 5 dB and 15 dB. For each SNR and a given resolution, 200 test examples are generated. For each example, first, f_1 is randomly selected and then we set $f_2 = f_1 + \Delta$, where Δ represents the desired resolution. With a limited sample size of $M = 50$, the Fourier-based periodogram method achieved a resolution threshold of $1/50 = 0.02$ Hz, whereas the full N-sample scenario yielded a resolution of $\Delta_f = 1/N = 0.0067$ Hz.

In Figs. 4(a) and (b), MSEs for various resolutions are shown for SNRs 5 dB and 15 dB, respectively. Starting from 50 samples, the learning-based approaches were able

TABLE I
PERFORMANCE COMPARISON OF HRSE AND PROPOSED TF-BASED METHOD FOR 15dB SNR

Method	Size	Resolution (NMSE in dB)				Relative time overhead
		0.5/150	1/150	2/150	3/150	
ESPRIT (50T)	-	0.14	-5.63	-41.39	-51.59	0.1X
ESPRIT (150T)	-	-52	**-65.33**	**-71.42**	**-73.21**	1X
TF dim=64, FF dim=256, N=2	160k	**-56.2**	-61.9	-65.83	-68.46	1.026X
TF dim=64, FF dim=256, N=1	110k	-50.26	-53.74	-55.98	-56	1.015X
TF dim=64, FF dim=128, N=2	110k	-52.26	-49.9	-57.84	-58.4	1.023X
TF dim=64, FF dim=512, N=2	259k	-54.21	-60.31	-62.1	-65.54	1.032X
TF dim=64, FF dim=1024, N=2	457k	-54.26	-59.34	-61.91	-64.2	1.042X
TF dim=32, FF dim=128, N=2	46k	-50	-46.23	-44.32	-50.95	1.015X

to have better resolution than *50T*. Among the learnable methods, *TF-50T+100P* was able to achieve the same (super)resolution ability as that of *150T*. Interestingly, at SNR 5 dB, the *CNN-50T+100P* and *DeepFreq* have lower errors compared to *150T* and *TF-50T+100P* for resolutions below $\Delta f = 1/150$. But for higher resolutions, the NMSEs for these methods are not decreasing as rapidly as *TF-50T+100P* and *150T*. Additionally, the CNN-LSTM model matches the performance of DeepFreq, indicating that these model architectures are inherently incapable of attaining super-resolution accuracy.

Next, we compared the methods for different SNRs while keeping the resolution as $\Delta_f = 1/150$, and the NMSEs are depicted in Figs. 4(c). For each SNR, 1000 test samples were used where the frequencies f_1 were picked randomly, and then we set $f_2 = f_1 + \Delta_f$.As expected, *50T* results in the highest error, while *150T* has approximately 50 dB less error at the cost of three times the number of measurements. In comparison, learning-based methods have significantly lower error than *50T* while using the same number of measurements. Specifically, $15 - 35$ dB lower for *CNN-50T+100P* and *DeepFreq*, and $45 - 50$ dB for *TF-50T+100P*.

D. Analysis

1) Computational Cost: We also evaluated the computational cost of our approach in comparison to traditional methods as shown in Table I. While our model incurs a 2.6% overhead with the same number of samples, this increased cost is compensated by a 66% reduction in sampling requirements. It is evident that the computational cost of traditional HRSE methods is the main source of latency and our proposed learning-based methods introduce negligible overhead. Overall, the proposed learning-based methods, especially the TF-50T+100P model, achieves super-resolution with significantly fewer measurements, even across various signal-to-noise ratios (SNRs).

2) Performance dependence on Model Scale: We now present a comprehensive evaluation of the Transformer-based model and the key design choices for accurate and efficient prediction of future samples. The experiment results for 15 dB SNR are reported in Table I, where TF dim is the transformer model (hidden) dimension, FF dim is the feedforward layer dimension, and N is the number of transformer layers, while the important insights are valid for other SNRs too. We observe that the feature dimension in the Transformer Encoder is the most important factor affecting prediction error, especially at higher resolutions. Additionally, for a fixed value of the feature dimension, the prediction error varies significantly with the ratio of the dimension of the intermediate linear layer in the MLP to the feature dimension. The best performance is achieved for the ratio of 4, and it degrades on both sides of this value. This is particularly important as this dimension determines the number of parameters in the MLP layer, which constitute the major part of the overall model size. We also note that there is substantial performance degradation for shallower models, suggesting that the depth of the Transformer Encoder is essential to generate high-quality representations, which can then be leveraged by the linear head to perform accurate predictions. These insights allow us to achieve super-resolution accuracy with small model sizes.

V. CONCLUSIONS

In this study, we present a novel learning-best approach for accomplishing super-resolution frequency estimation. Our approach involves training a model to predict future samples in a sum of complex exponentials using a limited sample set. By subsequently incorporating both the available and predicted samples, we successfully demonstrate the attainment of high-resolution estimation. This algorithm

proves valuable in scenarios where measurements come at a premium. Our ongoing efforts are directed towards expanding the method to predict samples from non-uniform datasets.

REFERENCES

[1] G. R. DeProny, "Essai experimental et analytique: Sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool, à différentes températures," *J. de l'Ecole polytechnique*, vol. 1, no. 2, pp. 24–76, 1795.

[2] J. A. Cadzow, "Signal enhancement — A composite property mapping algorithm," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. 36, pp. 49–62, Jan. 1988.

[3] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements," *Sampling Theory in Signal and Image Process.*, vol. 14, no. 1, pp. 17–47, 2015.

[4] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, 1997.

[5] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, vol. 8, 1983, pp. 336–339.

[6] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[7] B. D. Rao and K. V. S. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust., Speech and Signal Process.*, vol. 37, no. 12, pp. 1939–1949, 1989.

[8] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proc. IEEE*, vol. 74, no. 7, pp. 1044–1046, 1986.

[9] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech and Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[10] P. Stoica, J. Li, and H. He, "Spectral analysis of nonuniformly sampled data: A new approach versus the periodogram," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 843–858, 2009.

[11] P. Babu and P. Stoica, "Spectral analysis of nonuniformly sampled data – A review," *Digital Signal Process.*, vol. 20, no. 2, pp. 359–378, 2010.

[12] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE trans. info. theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[13] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, 2013.

[14] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3139–3153, 2015.

[15] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[16] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[17] G. Izacard, B. Bernstein, and C. Fernandez-Granda, "A learning-based framework for line-spectra super-resolution," in *Int. Conf. Acous., Speech and Signal Process. (ICASSP)*, 2019, pp. 3632–3636.

[18] Y. Guo, Z. Zhang, Y. Huang, and P. Zhang, "DOA estimation method based on cascaded neural network for two closely spaced sources," *IEEE Signal Process. Lett.*, vol. 27, pp. 570–574, 2020.

[19] G. Izacard, S. Mohan, and C. Fernandez-Granda, "Data-driven estimation of sinusoid frequencies," in *Advances Neural Info. Process. Syst. (NeurIPS)*, vol. 32, 2019.

[20] L. Wu, Z.-M. Liu, and Z.-T. Huang, "Deep convolution network for direction of arrival estimation with sparse prior," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1688–1692, 2019.

[21] A. M. Elbir, "DeepMUSIC: Multiple signal classification via deep learning," *IEEE Sensors Lett.*, vol. 4, no. 4, pp. 1–4, 2020.

[22] Y. Xie, M. B. Wakin, and G. Tang, "Data-driven parameter estimation of contaminated damped exponentials," in *Asilomar Conf. Signals Syst. Computers*, 2021, pp. 800–804.

[23] P. Pan, Y. Zhang, Z. Deng, and G. Wu, "Complex-valued frequency estimation network and its applications to superresolution of radar range profiles," *IEEE Trans. Geosci. Remote Sens*, vol. 60, pp. 1–12, 2022.

[24] P. Chen, Z. Chen, L. Liu, Y. Chen, and X. Wang, "Sdoa-net: An efficient deep-learning-based doa estimation network for imperfect array," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–12, 2024.

[25] M. Ali, A. A. Nugraha, and K. Nathwani, "Exploiting sparse recovery algorithms for semi-supervised training of deep neural networks for direction-of-arrival estimation," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2023, pp. 1–5.

[26] J. W. Smith and M. Torlak, "Frequency estimation using complex-valued shifted window transformer," *IEEE Geosci. Remote Sens. Lett.*, vol. 60, pp. 1–12, 2024.

[27] S. Biswas and A. C. Gurbuz, "Deep learning based high-resolution frequency estimation for sparse radar range profiles," in *IEEE Radar Conf. (RadarConf)*. IEEE, 2024, pp. 1–6.

[28] G. K. Papageorgiou, M. Sellathurai, and Y. C. Eldar, "Deep networks for direction-of-arrival estimation in low SNR," *IEEE Trans. Signal Process.*, vol. 69, pp. 3714–3729, 2021.

[29] Y. Jiang, H. Li, and M. Rangaswamy, "Deep learning denoising based line spectral estimation," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1573–1577, 2019.

[30] V. C. H. Leung, J.-J. Huang, Y. C. Eldar, and P. L. Dragotti, "Learning-based reconstruction of fri signals," *IEEE Tran. Signal Process.*, vol. 71, pp. 2564–2578, 2023.

[31] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[32] P. P. Vaidyanathan, *The theory of linear prediction*. Springer Nature, 2022.