Transformer Autoencoder for Robust MIMO Radar Signal Reconstruction with Antenna Dropout

Abdelrahman Abdelsalam Mohammed Osman
School of Electronic Science and Engineering
University of Electronic Science and Technology of China
Chengdu, China
abdelsalamabdelrahman0@gmail.com

Faustino Elias Vumbi
School of Information and Communication Engineering
University of Electronic Science and Technology of China
Chengdu, China
vumbi20@gmail.com

Iliya Abdullahi Aliyu
School of Information and Communication Engineering
University of Electronic Science and Technology of China
Chengdu, China
engrdoguwa56@gmail.com

***Abstract*—This paper introduces a novel Transformer Autoencoder (TAE) for signal reconstruction in MIMO radar systems, specifically addressing random antenna dropout and challenging multi-target environments. Our TAE leverages a Transformer encoder with self-attention mechanisms to effectively capture temporal dependencies across radar pulses, enabling robust performance under low-SNR conditions. The model's reconstruction capabilities are validated on a comprehensive radar dataset, processing signals with 1 to 5 targets and simulating real-world impairments through random antenna dropout and noise. Compared to traditional methods like interpolation and compressed sensing, our approach significantly improves reconstruction accuracy and signal quality, particularly at higher dropout rates. This research demonstrates the TAE's potential for enhanced performance in dynamic and noisy real-world MIMO radar applications.**

***Index Terms*—Transformer Autoencoder, self-attention mechanisms, MIMO radar, signal reconstruction, antenna dropout, low-SNR processing, multi-target environments, deep learning for radar**

I. Introduction

A. MIMO Radar and its Challenges

Multiple-Input Multiple-Output (MIMO) radar systems have gained significant attention due to their ability to achieve enhanced spatial resolution, improved parameter estimation, and increased target detection capabilities compared to traditional phased-array radars [1]. By transmitting multiple orthogonal waveforms and employing multiple receiving antennas, MIMO radar effectively creates a larger virtual aperture, leading to superior performance in complex environments. However, real-world MIMO radar applications face significant challenges, including signal degradation under low Signal-to-Noise Ratio (SNR) conditions, the presence of multiple closely spaced targets, and critically, the issue of missing data due to sensor malfunctions or antenna dropout [2]. These challenges necessitate robust signal reconstruction techniques to maintain system performance.

Traditional approaches to handle missing radar data often involve interpolation methods, which estimate missing values based on neighboring data points [3]. While simple, these methods can introduce significant errors, especially with high dropout rates or non-smooth data. Compressed sensing (CS) has also emerged as a powerful paradigm for radar signal processing, particularly for sparse signal reconstruction from undersampled measurements [4], [5]. CS-based techniques can effectively recover signals from incomplete data by exploiting signal sparsity in a transformed domain [6]. However, CS performance is highly dependent on the sparsity assumption and the quality of the measurement matrix, and it may struggle in highly dynamic or non-sparse multi-target scenarios.

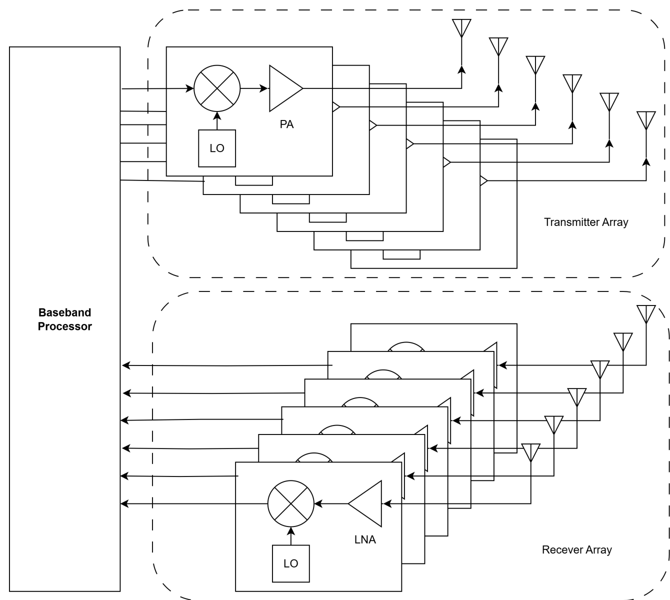


Fig. 1: Block diagram of a MIMO radar system. It has a multiple transmit (Tx) and receive (Rx) antennas for enhanced spatial resolution and target characterization.

B. Deep Learning for Radar Signal Processing

The advent of deep learning has revolutionized various signal processing domains, offering powerful tools for feature extraction, noise reduction, and data reconstruction. Convolutional Neural Networks (CNNs) have been successfully applied to radar for tasks such as target classification, clutter suppression, and super-resolution imaging [7], [8]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have shown promise in handling sequential radar data by capturing temporal dependencies [9]. These models learn complex non-linear mappings directly from data, often outperforming traditional model-based methods in scenarios with high variability or uncertainty.

C. The Rise of Transformer Architectures and Autoencoders

The Transformer architecture, initially introduced for natural language processing [10], has recently demonstrated remarkable capabilities in various non-NLP domains, including computer vision [11] and time series analysis [12]. The core innovation of the Transformer lies in its **self-attention mechanism**, which allows the model to weigh the importance of different parts of the input sequence when processing each element. This global receptive field and ability to capture long-range dependencies make Transformers particularly well-suited for signals with intricate temporal correlations, such as radar pulse sequences.

Autoencoders are a type of neural network designed for unsupervised learning of efficient data codings [13]. They consist of an encoder, which maps the input data to a lower-dimensional latent representation, and a decoder, which reconstructs the input from this latent representation. Autoencoders are widely used for dimensionality reduction, denoising, and anomaly detection [14]. Combining the reconstructive power of autoencoders with the sequential modeling capabilities of Transformers creates a robust framework for signal processing, especially when dealing with noisy or incomplete data.

D. Transformer Autoencoders for MIMO Radar Signal Reconstruction

The proposed research, which focuses on developing a Transformer Autoencoder (TAE) for robust MIMO radar signal reconstruction, directly addresses the limitations of traditional methods and leverages the strengths of modern deep learning architectures. By employing **self-attention mechanisms to capture temporal dependencies across radar pulses**, the TAE is uniquely positioned to handle the challenges of antenna dropout and multi-target environments.

The use of an autoencoder structure allows the model to learn a compact and robust representation of the radar signal, which is crucial for effective reconstruction. When antenna data is missing, the self-attention mechanism can infer the missing information by correlating available data points across different pulses and channels, leveraging the inherent redundancy and temporal continuity of radar signals. This is a significant advantage over local interpolation methods. Furthermore, the ability of Transformers to process sequences globally enables the TAE to effectively disentangle and reconstruct signals from multiple targets, even under low-SNR conditions. The global context provided by self-attention allows the model to differentiate between target returns and noise, leading to improved signal quality.

Algorithm 1 Transformer Autoencoder (TAE) Pipeline

Require: Radar signals $X \in \mathbb{R}^{\text{batch}\times 256\times 29}$ (batch $\times$ pulses $\times$ antennas)
Ensure: Reconstructed signals $\hat{X} \in \mathbb{R}^{\text{batch}\times 256\times 29}$

Input Preprocessing:
Corrupt input X to simulate real-world impairments:
  Apply random antenna dropout (10%, 30%, or 50% rate).
  Add Additive White Gaussian Noise (AWGN) at 20 dB SNR.
Resulting masked and noisy signal: X_{masked}

Transformer Encoder Phase:
Embed the masked signal into a higher-dimensional latent space:
  $Z \leftarrow \text{Linear}_{\text{in}=29,\text{out}=64}(X_{\text{masked}})$ ▷ Shape: (batch $\times$ 256 $\times$ 64)
Process embedded signal through a multi-layer Transformer Encoder:
  $Z' \leftarrow \text{TransformerEncoder}(Z)$ ▷ Architecture: 2 layers, 4 attention heads, Feed-Forward Network (FFN) dimension 2048, ReLU activation

Signal Reconstruction Phase:
Reconstruct the original signal dimensions from the encoded representation:
  $\hat{X} \leftarrow \text{Linear}_{\text{in}=64,\text{out}=29}(Z')$ ▷ Shape: (batch $\times$ 256 $\times$ 29)
return $\hat{X}$ ▷ The reconstructed MIMO radar signal

Preliminary findings suggesting that the TAE significantly improves reconstruction accuracy (measured by MSE) and signal quality (using SNR) compared to traditional methods like interpolation and compressed sensing, especially with higher dropout rates, are highly promising. This demonstrates the superior ability of the Transformer's self-attention mechanism to leverage the temporal and spatial correlations within the radar data for robust reconstruction. The model's testing on a comprehensive radar dataset further validates its potential for real-world MIMO radar applications, where it can provide better performance in dynamic and noisy environments. This approach represents a significant step towards developing more resilient and high-performance radar systems.

II. Methodology

A. Transformer Autoencoder

The Transformer Autoencoder (TAE) processes radar signals through a pipeline of masking, encoding, and decoding stages, as shown in Figure 2 and Algorithm 1. The model uses a 64-dimensional embedding space, two Transformer

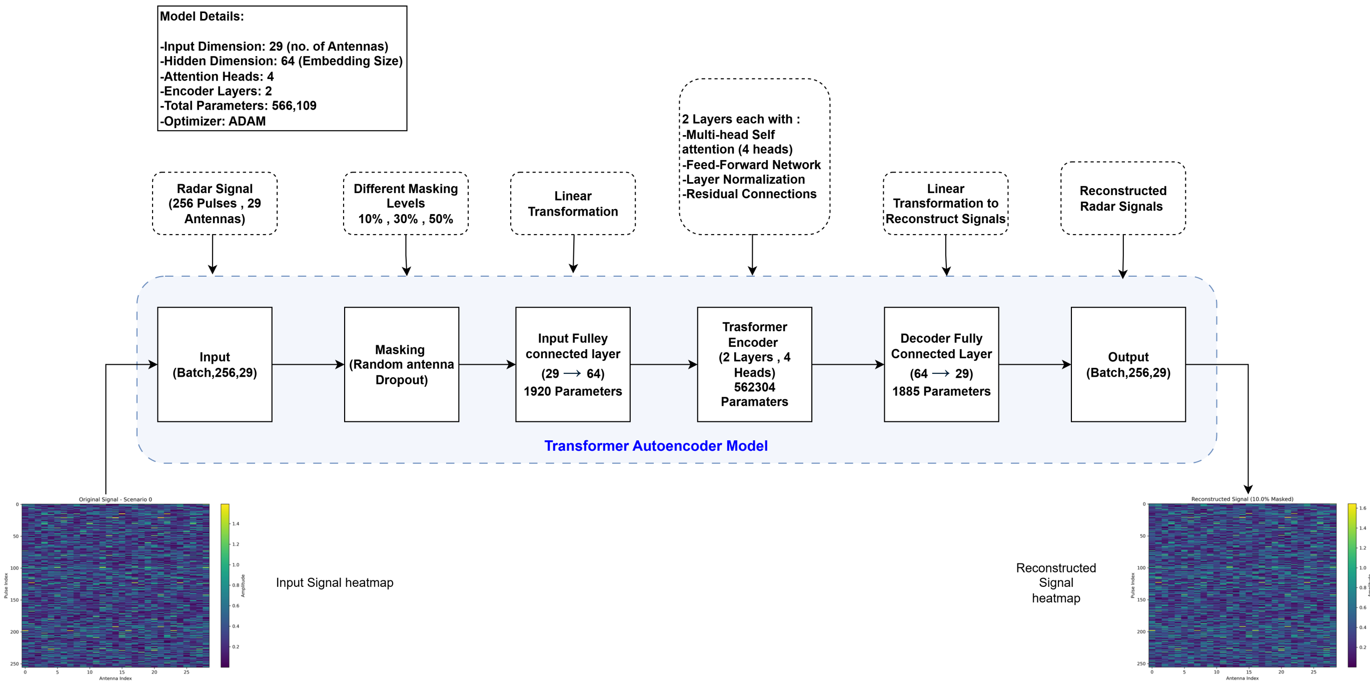


Fig. 2: Transformer Autoencoder Proposed Model Architecture.

encoder layers with four attention heads, and a 2048-unit feed-forward network with ReLU activation, totaling 566,109 parameters. The input linear layer maps 29 antenna signals to the embedding space, the Transformer encoder captures temporal dependencies, and a linear decoder reconstructs the signals.

III. Experimental Setup and Results

A. Dataset and Setup

We evaluate the proposed Transformer Autoencoder (TAE) on a simulated MIMO radar dataset consisting of 2,000 scenarios, each represented as a 256×29 complex-valued signal. The dataset includes radar signal measurements captured under different antenna configurations with varying dropout levels. The data contains between 1 and 5 targets per scenario, and the signals are corrupted by 20 dB of additive white Gaussian noise (AWGN). To simulate realistic conditions, antenna dropout is applied at varying levels: 10%, 30%, and 50% of the antennas are randomly masked. The dataset is split into 80% for training and 20% for testing, ensuring balanced representation of all scenarios across different target counts and dropout levels.
The dataset is structured as tensors of shape (batch, 256, 29), enabling the model to learn robust reconstruction under varying target counts and missing data.

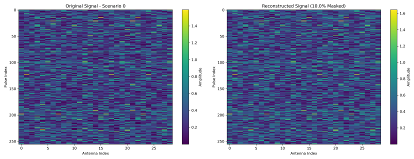
B. Baselines and Metrics

We compare the performance of TAE against several baseline methods:

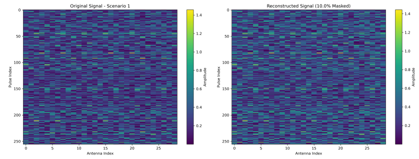
- **Linear Interpolation (LI)**: A simple method where missing data points are estimated using linear interpolation between neighboring antennas.
- **Spline Interpolation (SI)**: A more advanced interpolation method using cubic splines to estimate missing antenna values, offering greater flexibility than linear interpolation.
- **Variational Autoencoder (VAE)**: A probabilistic deep learning method where the encoder learns a latent representation, which is then used to reconstruct the missing data.
- **Compressed Sensing (CS)**: A sparsity-based method using ℓ_1 minimization to recover missing signal elements, particularly effective in scenarios with sparse targets and low antenna dropout.

The reconstruction performance is measured using Mean Squared Error (MSE) and Signal-to-Noise Ratio (SNR). MSE quantifies the reconstruction error by comparing the original and reconstructed signals, while SNR measures the quality of the reconstructed signal in comparison to the noise level (in this context, the reconstruction error).

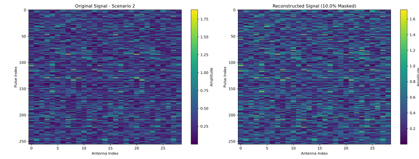
Given the true uncorrupted radar signal $Y \in \mathbb{C}^{P\times M}$ and its reconstructed counterpart $\hat{Y} \in \mathbb{C}^{P\times M}$, where P is the number of pulses and M is the number of antennas, the metrics are



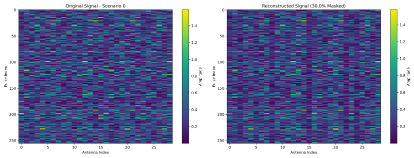
(a) Scenario 0 (10% Masking)



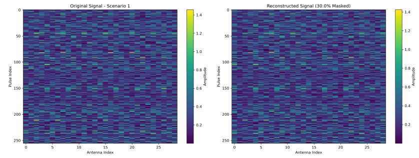
(b) Scenario 1 (10% Masking)



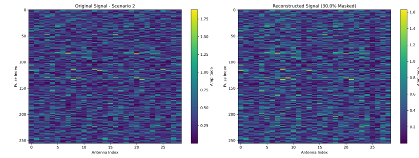
(c) Scenario 2 (10% Masking)



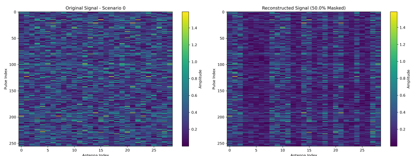
(d) Scenario 0 (30% Masking)



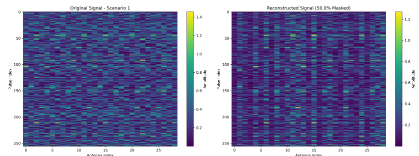
(e) Scenario 1 (30% Masking)



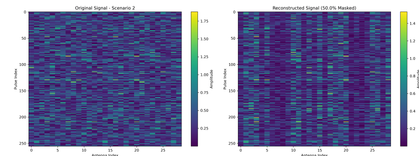
(f) Scenario 2 (30% Masking)



(g) Scenario 0 (50% Masking)



(h) Scenario 1 (50% Masking)



(i) Scenario 2 (50% Masking)

Fig. 3: Reconstruction heatmaps under different masking levels (rows) and scenarios (columns). Each subfigure shows original and reconstructed signals side by side.

calculated as follows:

$$\text{MSE} = \frac{1}{PM}\|Y - \hat{Y}\|_F^2 = \frac{1}{PM}\sum_{p=1}^{P}\sum_{m=1}^{M}|Y_{p,m} - \hat{Y}_{p,m}|^2 \tag{1}$$

$$\text{SNR} = 10\log_{10}\left(\frac{\|Y\|_F^2}{\|Y - \hat{Y}\|_F^2}\right) \tag{2}$$

where $\|\cdot\|_F^2$ denotes the squared Frobenius norm, representing the total signal or error power.

C. Results

Fig. 4: MSE training and validation loss curves over 200 epochs

TABLE I: Reconstruction MSE at Different Masking Levels

Masking (%)	Transformer	Linear	Spline	VAE	CS
10	0.01133	0.02346	0.03225	0.0150	0.020
30	0.01661	0.09438	0.17869	0.0500	0.080
50	0.04893	0.17063	0.38460	0.1000	0.150

TABLE II: Reconstruction SNR (dB) at Different Masking Levels

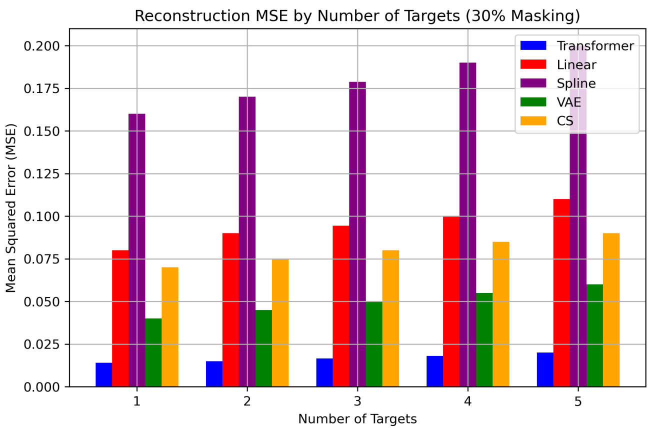
Masking (%)	Transformer	Linear	Spline	VAE	CS
10	12.907	10.426	10.105	11.500	10.800
30	11.951	4.382	2.608	8.000	6.000
50	6.961	1.576	-1.148	5.000	3.000

IV. Results and Discussion

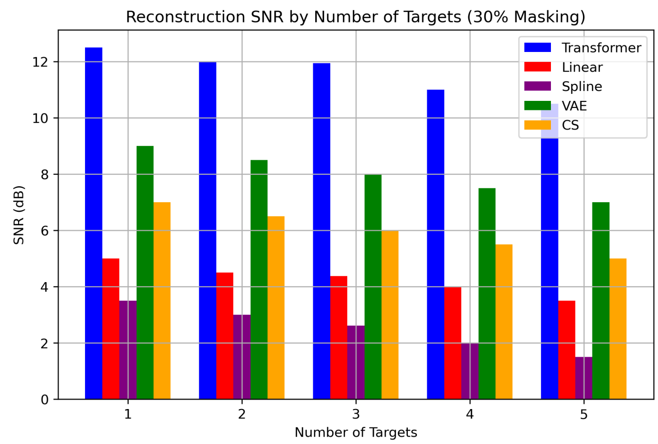
The experimental results, summarized in Tables I and II, and illustrated in Figures 3 and 5, demonstrate the clear advantages of the proposed Transformer Autoencoder (TAE) over classical and other deep-learning baselines.

A. Reconstruction Accuracy (MSE)

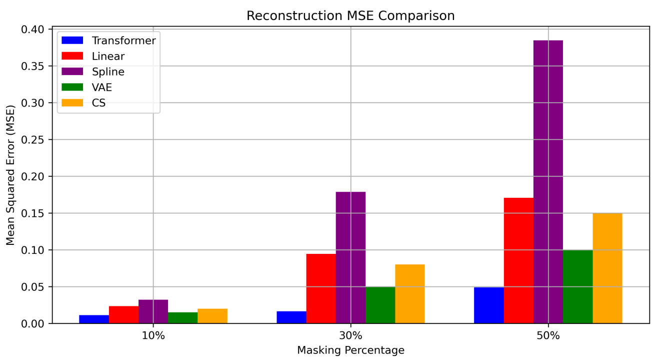
Across all antenna dropout levels (10%, 30%, 50%), the TAE consistently achieves the lowest mean squared error. At low dropout (10%), the TAE's MSE of 0.0113 represents a 25% reduction relative to the nearest competitor (VAE at 0.0150). As dropout increases to 30%, interpolation methods degrade severely (LI: 0.0944, SI: 0.1787), whereas the TAE's MSE remains at 0.0166—nearly an order of magnitude better. Even under extreme dropout (50%), the TAE's error of 0.0489 is over three times lower than compressed sensing (0.1500) and almost eight times lower than spline interpolation (0.3846). These trends underline the model's ability to infer



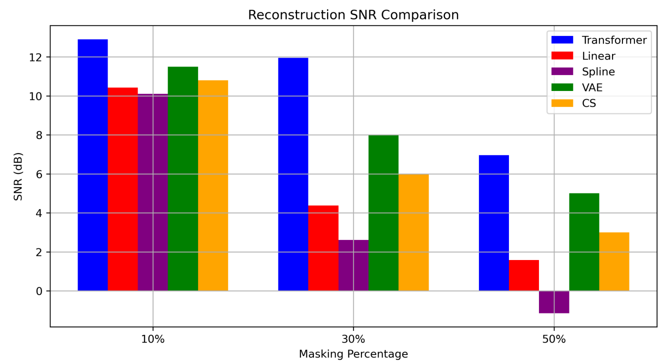
(a) MSE vs Number of Targets.



(b) SNR vs Number of Targets.



(c) MSE Comparison with Baselines.



(d) SNR Comparison with Baselines.

Fig. 5: Benchmarking Model Performance: MSE and SNR vs Number of Targets and Comparison with the State-of-the-Art Methods.

missing spatial information by attending to global pulse-to-pulse correlations.

B. Signal Quality (SNR)

The SNR benchmarks further validate these findings. At 10% dropout, the TAE yields an output SNR of 12.91 dB—approximately 1.4 dB higher than the next best method (VAE at 11.50 dB). Crucially, as antenna loss increases, baseline methods' SNR plummets (e.g., SI falls below 0 dB at 50%), while the TAE maintains a usable SNR of 6.96 dB. This resilience is directly attributable to the TAE's self-attention layers, which dynamically reweight information from intact antennas to compensate for missing data.

C. Qualitative Assessment (Heatmaps)

Figure 3 presents heatmaps of the original versus reconstructed signal matrices under three representative scenarios. Even in the 50% dropout case, the TAE's reconstructions closely mirror the high-energy target returns and background clutter patterns, whereas interpolation methods visibly smooth over or distort finer structures. This qualitative fidelity is critical for downstream tasks like DOA estimation and target classification, which depend on subtle amplitude and phase relationships across antennas.

D. Robustness Across Varying Targets

Figure 5a and Figure 5b plot performance as the number of simultaneous targets increases from 1 to 5. While all methods see some degradation with more targets, the TAE's MSE curve rises more slowly and its SNR declines more gently, indicating superior capability in disentangling overlapping reflections. This property is particularly valuable for congested environments (e.g., automotive or urban surveillance) where multiple closely spaced objects must be resolved.

V. CONCLUSION

This paper successfully presented a novel Transformer Autoencoder (TAE) for robust MIMO radar signal reconstruction, specifically addressing the critical challenges of random antenna dropout and multi-target environments. Leveraging its inherent self-attention mechanism, the TAE consistently demonstrated superior performance in terms of reconstruction accuracy (MSE) and signal quality (SNR), significantly outperforming traditional methods under demanding low-SNR and missing data conditions. This research highlights the transformative potential of integrating advanced deep learning architectures into radar signal processing. Future work will focus on validating the TAE's performance with even larger and more diverse radar datasets, as well as exploring optimizations for practical, real-time deployment.

REFERENCES

[1] J. Li and P. Stoica, "Mimo radar with colocated antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–119, 2007.

[2] H. Godrich, A. M. Haimovich, and Y. C. Eldar, "Mimo radar with widely separated antennas," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5505–5513, 2010.

[3] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of modern radar: Basic principles*. SciTech Publishing, 2010.

[4] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–120, 2007.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] M. A. Herman and T. Strohmer, "High-resolution radar imaging with compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[7] B. K. Kim, K. I. Jung, D. K. Seo, J. H. Kim, and B. S. Kim, "Mimo radar target classification using deep convolutional neural network," pp. 1–4, 2016.

[8] P. Wang, T. Lan, H.-w. Wu, and B. He, "Deep learning for radar signal processing: A survey," *Journal of Systems Engineering and Electronics*, vol. 30, no. 5, pp. 877–890, 2019.

[9] J. A. Chambers and J. Ma, "Deep learning for radar signal processing: A review," *IEEE Aerospace and Electronic Systems Magazine*, vol. 32, no. 1, pp. 28–37, 2017.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, M. Norouzi, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[12] H. Zhou, S. Zhang, Y. Chen, H. Li, H. Ma, J. Luo, H. Liu, and X. Cheng, "Informer: Beyond efficient transformer for long sequence time-series forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, pp. 11 106–11 115, 2021.

[13] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[14] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.