







Article

Deep Learning Framework Using Spatial Attention Mechanisms for Adaptable Angle Estimation Across Diverse Array Configurations [†]

Constantinos M. Mylonakis ^{1,‡} , Pantelis Velanas ^{2,‡} , Pavlos I. Lazaridis ^{3,‡} , Panagiotis Sarigiannidis ^{4,‡} ,
Sotirios K. Goudos ^{5,*}  and Zaharias D. Zaharis ^{1,‡} 

¹ School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; mylonakis@ece.auth.gr (C.M.M.); zaharis@auth.gr (Z.D.Z.)

² Acceligen Ltd., Nicosia 1066, Cyprus; pvelanas@acceligen.tech

³ School of Computing and Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK; p.lazaridis@hud.ac.uk

⁴ Informatics & Telecommunications Engineering, University of Western Macedonia, 50100 Kozani, Greece; psarigiannidis@uowm.gr

⁵ ELEDIA@AUTH, School of Physics Aristotle University of Thessaloniki, 54635 Thessaloniki, Greece

* Correspondence: sgoudo@physics.auth.gr

[†] This paper is an extended version of our paper published in 2024 13th International Conference on Modern Circuits and Systems Technologies (MOCASST), Sofia, Bulgaria, 26–28 June 2024; “3D Direction of Arrival Estimation: An Innovative Deep Neural Network Approach”, doi: 10.1109/MOCASST61810.2024.10615339.

[‡] These authors contributed equally to this work.

Abstract: Rapid advancement of wireless communication systems and the increasing need for accurate, real-time signal processing have driven innovations in direction-of-arrival (DoA) estimation techniques. This paper introduces a novel convolutional neural network (CNN) architecture that combines spatial attention mechanisms with a transfer learning framework to enhance both accuracy and versatility in DoA estimation. The model integrates spatial attention layers to dynamically prioritize signal regions with the highest information value, allowing it to isolate relevant signals and suppress interference in noisy or crowded signal environments. In addition, we utilize a transfer learning framework that enables the model to generalize across various antenna array configurations (i.e., planar, linear, and circular arrays) with minimal additional training. Extensive simulation results benchmark the proposed model against existing state-of-the-art methods for DoA estimation, achieving improved absolute error across diverse conditions. This hybrid approach not only enhances DoA estimation precision, but also significantly reduces retraining requirements when adapting to new array configurations, positioning it as a robust, scalable tool for next-generation wireless communication systems.

Keywords: direction-of-arrival (DoA) estimation; machine learning; convolutional neural networks; spatial attention; transfer learning; multiple-input multiple-output (MIMO) systems; wireless communications



Academic Editor: Pedro Antonio Gutiérrez

Received: 16 December 2024

Revised: 20 January 2025

Accepted: 22 January 2025

Published: 24 January 2025

Citation: Mylonakis, C.M.; Velanas, P.; Lazaridis, P.I.; Sarigiannidis, P.; Goudos, S.K.; Zaharis, Z.D. Deep Learning Framework Using Spatial Attention Mechanisms for Adaptable Angle Estimation Across Diverse Array Configurations. *Technologies* **2025**, *13*, 46. <https://doi.org/10.3390/technologies13020046>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rollout of 5G and the anticipated development of 6G networks have introduced new demands for high-precision, real-time direction-of-arrival (DoA) estimation. Core applications, including beamforming, interference management, and spatial signal processing, rely on accurate DoA estimation to optimize network performance and maximize data throughput. As user density and data traffic increases, precise DoA estimation becomes

critical to minimizing latency and reducing signal interference in congested environments. Emerging use cases, for example autonomous vehicles, radar and sonar applications, augmented reality, and industrial IoT, require robust DoA solutions to ensure seamless and reliable connectivity in dynamic and complex settings. To meet these demands, advanced DoA estimation methods must address challenges including high mobility, multipath effects, and varying propagation conditions, which are increasingly prevalent in modern wireless systems.

1.1. Literature Review

DoA estimation has been a foundational technique within wireless communications, radar, sonar, and sensor networks [1,2]. Initial research into DoA primarily leveraged signal processing algorithms, including conventional beamforming [3], Capon's minimum variance method [4], multiple signal classification (MUSIC) [5], and estimation of signal parameters via rotational invariance techniques (ESPRIT) [6]. Although these pioneering methods laid the foundation for array processing, they revealed inherent limitations in complex environments, especially when managing multiple source signals with low signal-to-noise ratios (SNR). In addition, these classical methods encounter limitations in high-interference or multi-signal scenarios, and are typically tailored to fixed antenna arrays, which constrains their adaptability in the increasingly dynamic configurations required by modern networks. Subspace methods remain computationally intensive, and require highly controlled conditions for accurate performance, limiting their suitability for real-time applications or scenarios with non-linear interference patterns [7].

With the evolution of signal processing techniques, statistical and model-based methods emerged, further advancing the robustness and adaptability of DoA estimation. Bayesian estimation techniques introduced probabilistic frameworks that addressed uncertainties, enhancing robustness in complex signal conditions [8]. Maximum likelihood methods, particularly those coupled with iterative refinement algorithms, e.g., expectation maximization, gained popularity for achieving high-resolution DoA estimations [9,10]. Deep research has also been applied on heuristic and adaptive methods to alleviate computational constraints [11]. Although these methods surpass traditional algorithms in accuracy, their computational demands make them less suitable for real-time adaptation in dynamic signal environments. This evolution underscored the increasing need for more flexible and efficient DoA estimation techniques to support next-generation network requirements.

The advent of machine learning (ML) marked a shift in DoA research [12–16], as traditional algorithms began to be augmented or replaced by data-driven approaches. Early ML models, i.e., support vector machines and k-nearest neighbors, demonstrated that supervised learning could help DoA systems generalize better across varying signal conditions. Nevertheless, these techniques could not fully capture the complex, high-dimensional relationships inherent in real-world signals, which limited their efficacy in challenging scenarios involving multi-path and dense interference. Recent advancements in deep learning [17] have introduced more powerful architectures, including feed-forward neural networks (FFNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), which have shown significant promise for DoA estimation in complex environments. FFNNs have been applied for their simplicity and computational efficiency, particularly in scenarios where low-latency DoA estimation is critical [18]. Early CNN-based models were largely designed for single-source DoA estimation and exhibited reduced efficacy in high-noise, multi-source environments [19,20]. However, CNNs have been used mainly to exploit spatial features in signal data, allowing for robust DoA estimation under multi-signal and multi-path conditions [21,22]. RNN-based approaches have

also been explored to model temporal dependencies within time-sequenced signals, which can improve accuracy in rapidly changing signal environments [23].

Recent endeavors in DoA estimation research have explored innovative array configurations, algorithmic improvements, and novel hardware setups, leading to significant enhancements in localization accuracy and applicability across various domains. Ref. [24] introduced a 2-D DoA estimation framework leveraging an L-shaped array, which effectively enhances angular resolution and spatial diversity. Ref. [25] investigated the use of a partially connected receiving reconfigurable intelligent surface for 3D localization, providing comprehensive positioning error analysis and a tailored algorithmic design. Ref. [26] explored virtual antennas within the MUSIC algorithm framework, demonstrating their capability to improve spectral resolution and target separation in constrained physical arrays. This innovation broadens the applicability of subspace methods in real-time applications and environments with limited sensor resources. Additionally, Ref. [27] showcased a cost-effective DOA estimation system using a single antenna and low-cost backscattering tags. Their approach reduces hardware complexity while achieving competitive performance, making it suitable for IoT applications and resource-constrained environments. Ref. [28] proposed an innovative generative approach to enhance the robustness of DoA estimation under jamming attacks. By leveraging a generative framework, this method effectively reconstructs corrupted signal data, mitigating the adverse effects of intentional interference. This advancement is particularly impactful in secure wireless communication scenarios, where resilience against adversarial attacks is crucial for maintaining accurate signal localization.

Hybrid models combining CNNs and RNNs have been proposed to simultaneously capture spatial and temporal features, yielding promising results for DoA estimation in mixed-line-of-sight and non-line-of-sight scenarios [29]. Recent work has introduced transformer-based architectures, which leverage attention mechanisms to further enhance DoA estimation by focusing on the most informative parts of the signal data [30]. Combining CNNs with advanced attention mechanisms allows models to selectively focus on specific regions within the input signal and improve accuracy in high-interference environments [31,32]. Attention-based models, i.e., transformers, have shown promise in DoA applications, as they enable the network to weigh different input regions dynamically, leading to more precise signal localization under adverse conditions [33,34]. While attention mechanisms are increasingly popular in various domains, their application in spatial DoA tasks is still novel.

Another crucial challenge in DoA estimation is the model adaptability to different antenna configurations. Most deep learning models are trained for specific array structures (e.g., uniform planar or linear arrays), and this specificity often requires retraining or fine-tuning when applied to new configurations. Transfer learning has emerged as a promising solution to this limitation, allowing models to be pre-trained on one configuration and adapted to others with minimal additional data [35–37]. This approach not only reduces computational demands, but also improves generalization across varied settings. Recent research has demonstrated the effectiveness of transfer learning in tasks that share similar spatial structures, suggesting that DoA models can be enhanced to handle diverse array types without requiring extensive retraining.

1.2. Motivation and Contributions

Recent research efforts focus on addressing the limitations of deep learning models in DoA estimation. Conventional and even deep learning-based models tend to be specialized for specific antenna types or configurations. However, in practical applications, the requirements of the system demand flexibility to work with multiple configurations

and accommodate dynamic, high-noise environments. This is where spatial attention and transfer learning prove highly valuable: spatial attention allows the model to selectively focus on relevant parts of the input, improving its ability to isolate signals of interest even when interference is high. Transfer learning provides the model with the capability to adapt to various antenna array geometries without requiring extensive data retraining for each configuration.

Given the limitations of existing DoA estimation methods, this study is motivated by the need to develop a novel CNN encompassing spatial attention layers, capable of robust and improved performance in challenging signal environments and adaptability across diverse antenna array configurations. The primary contributions of this paper include the following:

- The introduction of a novel CNN architecture that encompasses spatial attention layers, tailored for enhanced accuracy on DoA estimations and robustness.
- The development of a transfer learning framework that enables the model to generalize across different antenna array configurations with minimal retraining.
- The analysis of the model's complexity before and after adding the attention layer focuses on the performance trade-offs. Transfer learning complexity and training times are also evaluated to assess adaptation efficiency.
- Comprehensive simulation results demonstrating the model's superior performance, achieving minimal error in various noise levels and array types, indicating its robustness and versatility.

The proposed method holds significant potential across various industries, particularly in telecommunications, radar systems, and autonomous vehicles. In telecommunications, precise DoA estimation is crucial for optimizing beamforming in 5G and future 6G networks, enhancing signal reception while minimizing interference. The ability of our model to generalize across different antenna configurations ensures robust performance in heterogeneous network environments, enabling more efficient resource allocation in massive MIMO systems. In radar and autonomous systems, accurate DoA estimation is essential for situational awareness and target tracking. Our approach enhances the reliability of phased-array radar systems by efficiently adapting to varying sensor geometries, which is particularly beneficial in defense, aerospace, and weather monitoring applications. In autonomous vehicles, precise localization of signal sources, such as GPS, V2X communication, and obstacle detection sensors, is critical for safe navigation.

1.3. Structure

The remainder of this paper is structured as follows. Section 2 outlines the signal and antenna array models across various configurations. Section 3 details the experimental setup, simulation parameters, and datasets used. Section 4 describes the proposed model architecture, emphasizing the spatial attention mechanism and the transfer learning framework. Section 5 analyzes the results, highlighting the model's performance and superiority compared to existing methods. Finally, Section 6 provides conclusions and discusses potential future research directions.

2. Signal Model and Array Geometry Analysis

The arrival of each incoming signal can be described by an angle pair (θ_k, ϕ_k) , where $k = 1, \dots, K$. Alternatively, it can be represented by a radial unit vector in the spherical coordinate system as

$$\mathbf{v}_k = \cos \phi_k \sin \theta_k \mathbf{x}_0 + \sin \phi_k \sin \theta_k \mathbf{y}_0 + \cos \theta_k \mathbf{z}_0, \quad (1)$$

where \mathbf{x}_0 , \mathbf{y}_0 , and \mathbf{z}_0 denote unit vectors along the x , y , and z axes, respectively. For the wavefront of the k -th signal, there is a time delay required to travel from the position of the m -th array element to the origin of the coordinate system. This time delay is given by

$$\tau_m(\theta_k, \phi_k) = \frac{\mathbf{r}_m \cdot \mathbf{v}_k}{c}, \quad m = 1, \dots, M, \quad (2)$$

where c represents the speed of light, M is the total number of antenna elements, and \mathbf{r}_m is the position vector of the m -th array element. These time delays represent the phase differences of the incoming signals as received by the array elements. Collectively, these phase differences are expressed through the so-called “steering vector”.

Different array geometries provide unique spatial characteristics that influence DoA estimation performance. This section presents a comprehensive analysis of these geometries, focusing on the signal models, the construction of their steering vectors, and spatial properties relevant to DoA estimation in complex signal environments. Figure 1 illustrates three distinct antenna array configurations: the uniform planar array (UPA), uniform linear array (ULA), and uniform circular array (UCA). Each configuration is shown capturing signals from a transmitter, highlighting their spatial arrangement and how they differ in terms of layout.

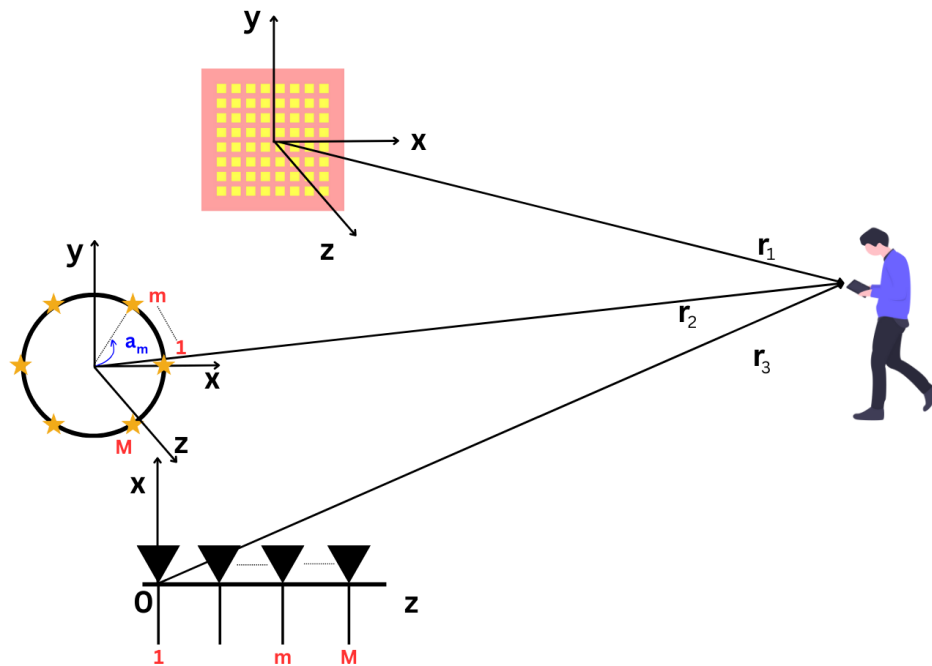


Figure 1. Schematic of three antenna array configurations—UPA, ULA, and UCA—capturing signals from a transmitter.

2.1. Uniform Linear Array

A ULA consists of M antenna elements aligned linearly with equal spacing d between consecutive elements. Given K narrowband plane waves impinging on the ULA at azimuth angles θ_k (for $k = 1, 2, \dots, K$), the received signals at the array can be expressed as

$$\mathbf{y}(t) = \sum_{k=1}^K \mathbf{a}(\theta_k) s_k(t) + \mathbf{n}(t), \quad (3)$$

where $\mathbf{y}(t) \in \mathbb{C}^{M \times 1}$ is the received signal vector at time t , $s_k(t)$ is the signal from the k -th source, $\mathbf{n}(t) \in \mathbb{C}^{M \times 1}$ is the additive noise vector, and $\mathbf{a}(\theta_k) \in \mathbb{C}^{M \times 1}$ is the steering vector for the k -th source at angle θ_k .

The steering vector $\mathbf{a}(\theta_k)$ for the ULA is formulated based on the phase differences introduced by the array structure. For a ULA with inter-element spacing d and wave number $\beta = \frac{2\pi}{\lambda}$, the steering vector can be written as

$$\mathbf{a}(\theta_k) = \begin{bmatrix} 1 \\ e^{j\beta d \cos \theta_k} \\ \vdots \\ e^{j(M-1)\beta d \cos \theta_k} \end{bmatrix}. \quad (4)$$

This formulation captures the progressive phase shifts across the array elements as a function of the incoming signal's polar angle. The ULA's primary advantage lies in its simplicity and strong directional sensitivity along the array's axis, making it suitable for DoA estimation with respect to polar angles.

2.2. Uniform Planar Array

A UPA is a two-dimensional array structure with $M_1 \times M_2$ elements arranged in a grid pattern, typically spaced d units apart along both the x - and y -axes. This configuration allows for both azimuth and polar DoA estimation, providing improved spatial resolution. Consider a UPA receiving K narrowband plane waves from sources located at polar angles θ_k and azimuth angles ϕ_k . The received signal model is given by

$$\mathbf{y}(t) = \sum_{k=1}^K \mathbf{a}(\theta_k, \phi_k) s_k(t) + \mathbf{n}(t), \quad (5)$$

where $\mathbf{a}(\theta_k, \phi_k) \in \mathbb{C}^{M_1 M_2 \times 1}$ is the steering vector for a source at angle (θ_k, ϕ_k) . For a source with polar θ_k and azimuth ϕ_k , the UPA steering vector can be decomposed as a Kronecker product of two steering vectors along the x - and y -axes,

$$\mathbf{a}(\theta_k, \phi_k) = \mathbf{a}_x(\theta_k, \phi_k) \otimes \mathbf{a}_y(\theta_k, \phi_k), \quad (6)$$

where \otimes denotes the Kronecker product. Vectors $\mathbf{a}_x(\theta_k, \phi_k)$ and $\mathbf{a}_y(\theta_k, \phi_k)$ are defined as

$$\mathbf{a}_x(\theta_k, \phi_k) = \begin{bmatrix} 1 \\ e^{jkd \cos \phi_k \sin \theta_k} \\ \vdots \\ e^{j(M_1-1)kd \cos \phi_k \sin \theta_k} \end{bmatrix}, \quad \mathbf{a}_y(\theta_k, \phi_k) = \begin{bmatrix} 1 \\ e^{jkd \sin \phi_k \sin \theta_k} \\ \vdots \\ e^{j(M_2-1)kd \sin \phi_k \sin \theta_k} \end{bmatrix}. \quad (7)$$

2.3. Uniform Circular Array

A UCA consists of M elements arranged in a circular formation with radius R , providing 360-degree coverage in the azimuth plane. This array type is particularly effective for omnidirectional DoA estimation. Assuming a signal arriving at polar θ_k and azimuth ϕ_k , the steering vector for a UCA is defined by the spatial phase shift relative to each element. The steering vector $\mathbf{a}(\theta_k, \phi_k)$ for the UCA is given by

$$\mathbf{a}(\theta_k, \phi_k) = \begin{bmatrix} e^{j\beta R \cos(\phi_k - \alpha_1) \sin \theta_k} \\ e^{j\beta R \cos(\phi_k - \alpha_2) \sin \theta_k} \\ \vdots \\ e^{j\beta R \cos(\phi_k - \alpha_M) \sin \theta_k} \end{bmatrix}, \quad (8)$$

where $\alpha_m = \frac{2\pi(m-1)}{M}$ is the angular position of the m -th antenna element.

2.4. Covariance Matrix

The covariance matrix, denoted as \mathbf{R}_{xx} , is a critical tool in analyzing the properties of the received signals, as it encapsulates the spatial correlations between the different received signals, and is defined as

$$\mathbf{R}_{xx} = \mathbb{E}[\mathbf{y}(t)\mathbf{y}^H(t)]. \quad (9)$$

Substituting the received signal model into this equation leads to

$$\mathbf{R}_{xx} = \sum_{k=1}^K \mathbf{a}(\theta_k)\mathbb{E}[s_k(t)s_k^H(t)]\mathbf{a}^H(\theta_k, \phi_k) + \sigma^2\mathbf{I}. \quad (10)$$

This equation can succinctly be expressed as

$$\mathbf{R}_{xx} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^H + \sigma^2\mathbf{I}, \quad (11)$$

where \mathbf{A} is the steering matrix containing the steering vectors for each source and \mathbf{R}_{ss} is the covariance matrix of the source signals, representing the correlation between the individual source signals $s_k(t)$. Each element of \mathbf{R}_{ss} captures the expected value of the product between signals from different sources, $\mathbb{E}[s_k(t)s_l^H(t)]$, for all k and l .

\mathbf{R}_{xx} captures the information needed for various conventional DoA estimation algorithms, e.g., MUSIC and ESPRIT, which exploit the eigenstructure of \mathbf{R}_{xx} to identify the directions of the incoming signals. Differences in the configuration of the array significantly affect the structure of the covariance matrix. In a UPA, the covariance matrix incorporates additional dimensions, reflecting the influence of the polar angle on the received signal. This added complexity allows for more accurate estimations in scenarios with multiple signal paths. Similarly, for a UCA, the covariance matrix reflects the circular symmetry of the array, which can enhance the resolution of DoA estimates, especially in scenarios with closely spaced sources.

3. Experimental Setup

In our analysis, we construct a diverse dataset that includes data from various antenna configurations: UPA, ULA, and UCA. This dataset serves as the foundational input for the proposed CNN model. Given the intricacies and demands of DoA estimation, careful attention is paid to dataset creation to ensure robustness and reliability. In this section, we detail the process of dataset design, emphasizing the considerations for representing different array configurations. Additionally, we outline the dataset's structure and the performance metrics employed to evaluate the model, ensuring a comprehensive framework for analysis and validation.

3.1. Dataset and Signal Generation

Our dataset comprises 10^6 records of incoming signal data, offering a robust foundation for the CNN to learn and generalize from multiple DoA scenarios. These records represent a broad spectrum of AoA situations, simulating different signal sources and their respective arrival angles. The dataset is then divided into two parts: 90% is allocated for training the CNN, while the remaining 10% is set aside for validation purposes. This split allows the model to learn effectively while preserving a subset for unbiased performance evaluation.

For each record, the azimuth angle (ϕ) ranges between 0° and 360° , while the polar angle (θ) ranges between 0° and 60° . The dataset includes three variations based on

the array type, which directly impacts the dimensions and complexity of the covariance matrices generated. In our analysis,

1. UPA is a 4×4 array, resulting in a total of 16 elements. For each record associated with the UPA configuration, we generate a 16×16 covariance matrix \mathbf{R}_{xx} . Since this is a complex matrix, we decompose it into three parts for CNN compatibility: real, imaginary, and phase components. This yields a $10^6 \times 16 \times 16 \times 3$ matrix, where each record has dimensions $16 \times 16 \times 3$. This ensures that both the amplitude and phase information, crucial for accurate signal localization, are preserved.
2. ULA is a linear arrangement of 8 antenna elements in a single row. For each record, this results in a 4×4 covariance matrix \mathbf{R}_{xx} . Similar to the UPA, this matrix contains complex elements, so we decompose it into real, imaginary, and phase components. Each record for the ULA configuration is thus represented as a $10^6 \times 8 \times 8 \times 3$ matrix. By including this configuration, the dataset allows the CNN to learn from a simpler array structure with fewer elements, which is useful for applications where space or cost constraints limit the number of antennas.
3. UCA consists of 8 elements arranged in a circular pattern, with a radius of 0.5 wavelengths. Similarly to the ULA, the UCA produces a 8×8 covariance matrix \mathbf{R}_{xx} for each record, which is also decomposed into real, imaginary, and phase parts. The resulting data for each UCA record are thus represented as a $10^6 \times 4 \times 4 \times 3$ matrix. The circular arrangement introduces unique spatial properties, e.g., uniform sensitivity across azimuth angles, making it particularly suitable for applications requiring omnidirectional reception.

Each sample of our dataset is created with a specified SNR value, varying between -20 to 20 dB. The variation in SNR values is critical for ensuring that the model learns to distinguish signal patterns under both high and low signal quality conditions. Alongside the covariance matrix data, the dataset includes an angle label matrix with dimensions $S \times K$, where $S = 10^6$ (total records) and K represents the total number of distinct angle sets. This matrix contains the actual AoA values (θ_k, ϕ_k) for each record, providing a basis for comparing the CNN's predictions with the true AoAs.

3.2. Performance Evaluation Metrics

To rigorously evaluate the model's performance in DoA estimation, we employ a suite of performance metrics focusing on angular error and computational efficiency. These metrics include mean absolute error (MAE) and root mean squared error (RMSE) in the angle domain, along with inference time, to provide a comprehensive understanding of both accuracy and speed in predictions.

MAE is the primary metric used to assess the accuracy of the model in estimating angles. In the angle domain, MAE captures the absolute deviation between the estimated angles and the true angles across multiple sources. MAE is defined as

$$\text{MAE} = \frac{1}{Q} \sum_{q=1}^Q (|\hat{\theta}_k - \theta_k| + |\hat{\phi}_k - \phi_k|), \quad (12)$$

where Q represents the total number of sources in each instance for the considered scenario, $\hat{\theta}_k$ and $\hat{\phi}_k$ denote the estimated polar and azimuth angles, respectively, for the k -th source, and θ_k and ϕ_k are the true polar and azimuth angles, respectively, for the k -th source.

To gain further insight into the model's error distribution, we also calculate RMSE. Unlike MAE, RMSE penalizes larger errors more heavily, offering a deeper view of the model's performance consistency across instances. RMSE is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{Q} \sum_{q=1}^Q ((\hat{\theta}_k - \theta_k)^2 + (\hat{\phi}_k - \phi_k)^2)}, \quad (13)$$

where the terms are defined similarly to those in the MAE calculation. In this equation, $(\hat{\theta}_k - \theta_k)^2$ and $(\hat{\phi}_k - \phi_k)^2$ represent the squared differences of predicted and actual polar and azimuth angles, respectively.

To complement the analysis of RMSE and MAE, we also evaluate the mean absolute percentage error (MAPE) to assess the relative error as a percentage of the true values. MAPE is insightful when the scale of the target values varies significantly, as it standardizes the error by dividing the absolute error by the magnitude of the true value. This metric is defined as

$$\text{MAPE} = \frac{1}{Q} \sum_{q=1}^Q \left(\frac{|\hat{\theta}_k - \theta_k|}{\theta_k} + \frac{|\hat{\phi}_k - \phi_k|}{\phi_k} \right) \times 100. \quad (14)$$

Another critical performance aspect is the inference time per instance, which reflects the model's real-time processing capability. For practical DoA applications, low-latency estimation is essential, particularly in systems that require rapid response times, such as adaptive beamforming or dynamic interference management. We assess the model's inference time under two scenarios:

1. Batch processing: the time required to process a batch of instances simultaneously, which evaluates the model's efficiency in high-throughput scenarios.
2. Single-instance processing: the time taken to process a single instance, providing insights into the model's feasibility for real-time or near-real-time applications.

4. Proposed CNN with Spatial Attention Mechanisms Architecture

4.1. Hyperparameter Tuning

The hyperparameter tuning process for our CNN model with spatial attention and transfer learning is a crucial step in optimizing the model's accuracy, robustness, and adaptability in different antenna array configurations. Given the complexity of the model, tuning is performed iteratively, focusing on parameters that govern the feature extraction layers, attention layers, and transfer learning components. Each hyperparameter is evaluated for its impact on both convergence speed and final performance metrics.

The model comprises multiple convolutional layers designed to extract spatial features from the covariance matrices. Filter size and number of filters are two of the most impactful hyperparameters. We experiment with filter sizes ranging from 3×3 to 7×7 , finding that higher filter sizes captured localized spatial features effectively, which is essential in high-noise environments. Increasing the number of filters per layer, starting with 64 and progressing to 128 and 256 in subsequent layers, enhances performance, particularly in deeper layers. However, beyond a threshold, larger filter counts increased computational cost without significant performance gains. Ultimately, using a combination of high filter sizes (7×7) and gradually increasing the filter count as the network deepened resulted in the best trade-off between computational efficiency and accuracy.

Key hyperparameters for tuning the spatial attention include the attention pooling size and regularization factor. For the pooling size, a 2×2 configuration is found to balance information retention and computation overhead effectively. The regularization factor, which controls the sensitivity of the attention map to specific regions, is tuned to prevent

overfitting on noise. Overly strong regularization dampens the model's ability to discern relevant features, while too weak regularization results in overfitting on high-noise samples. After evaluating several values, we found an optimal regularization factor that allows the model to generalize well across diverse SNR conditions and antenna configurations.

Transfer learning is a central feature of the model, allowing it to adapt to different antenna array configurations with minimal retraining. The primary parameters tuned are learning rate adjustments and layer freezing configurations. Given the unique spatial characteristics of each array configuration, we experiment with freezing different layers in the convolutional backbone. Freezing too many layers reduced the model's adaptability to new configurations, while freezing too few increased training time. A balanced approach involved freezing only the initial convolutional layers while allowing deeper layers, including the attention and fully connected layers, to fine-tune to each configuration. This approach achieves efficient transfer without significant re-training, especially when adapting from one array type to another (e.g., ULA to UPA). Additionally, adjusting the learning rate for the fine-tuning phase is critical; a low learning rate (i.e., 10^{-3}) provides stable convergence without diverging from the pre-trained weights.

The choice of optimizer and regularization techniques further influence the model's performance. We experiment with standard optimizers, including Adam and SGD with momentum, both offering unique advantages. Adam provides rapid convergence, which is crucial for initial model evaluation, though SGD with momentum proves more effective during the fine-tuning phase, helping the model generalize better across configurations. Validating on that trade-off, Adam proves the most suitable for our architecture, providing higher accuracy. For regularization, dropout layers are incorporated into the fully connected layers, and batch normalization is applied after each convolutional layer to mitigate overfitting and stabilize training. Different dropout rates (0.2 to 0.5) were tested, with a moderate rate of 0.3 offering the best trade-off between training speed and generalization.

4.2. Proposed Architecture

The proposed model's input is the covariance matrix \mathbf{R}_{xx} . The CNN initiates with three convolutional layers designed to progressively extract spatial features from the covariance matrix. Early layers employ higher filter sizes of 7×7 , with high depth (256 filters), enabling the network to capture local, fine-grained spatial features essential for resolving closely spaced DoAs. Each convolutional layer is followed by a ReLU activation function, $f(x) = \max(0, x)$, which introduces non-linearity, allowing the model to represent complex patterns. Moreover, appropriate strides and padding are used to preserve spatial resolution while ensuring that the convolutional operations encompass the full extent of the input matrix. The max-pooling layer is interspersed between the second and the third convolutional layer, to down-sample the spatial dimensions and reduce computational demands. Each pooling operation halves the spatial resolution, preserving essential information by retaining the most prominent features from local regions.

After the core feature extraction layers, spatial attention mechanisms are introduced. These attention layers are integrated to enhance the model's focus on high-information regions of the input, particularly beneficial in noisy or multi-source scenarios. The attention mechanism integrated into our neural network operates as a channel-wise attention layer designed to enhance the model's ability to focus on the most relevant features for accurate direction-of-arrival (DoA) estimation. The attention layer is implemented using a two-step process:

1. A global average pooling (GAP) operation is applied to the feature map \mathbf{F} , where $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$ represents the output of the convolutional layers. Here: H and W are

the spatial dimensions of the feature map and C is the number of channels (filters) in the feature map.

The GAP operation collapses the spatial dimensions, producing a vector $\mathbf{z} \in \mathbb{R}^C$,

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_{ijc}, \quad (15)$$

where F_{ijc} is the value of the feature map at spatial location (i, j) for channel c . This vector \mathbf{z} captures the global representation of each channel's contribution. GAP aggregates spatial information into a single value for each channel, providing a global summary of each filter's activation. This is critical for capturing spatially invariant features relevant to the DoA task.

2. The pooled vector \mathbf{z} is passed through a two-layer feed-forward network to compute attention weights for each channel,

$$\mathbf{w} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{z})), \quad (16)$$

where $\mathbf{W}_1 \in \mathbb{R}^{C \times (C/r)}$ and $\mathbf{W}_2 \in \mathbb{R}^{(C/r) \times C}$ are weight matrices of the fully connected layers. Here, r is a reduction ratio (e.g., $r = 2$), which reduces the parameter complexity while preserving key information, and σ is the sigmoid activation function, ensuring the weights \mathbf{w} lie in the range $[0, 1]$.

These weights $\mathbf{w} \in \mathbb{R}^C$ signify the importance of each channel. The two-layer feed-forward network acts as a lightweight attention mechanism that captures inter-channel dependencies. By using a non-linear ReLU activation followed by a sigmoid, the network learns complex, non-linear relationships between channels and maps them to importance weights.

3. The weights \mathbf{w} are applied to the original feature map \mathbf{F} through channel-wise multiplication,

$$\hat{\mathbf{F}}_{ijc} = w_c F_{ijc}, \quad (17)$$

where $\hat{\mathbf{F}}_{ijc}$ is the recalibrated feature map. This operation amplifies the channels with higher weights and suppresses less relevant ones. By multiplying the original feature maps by the learned attention weights, the model dynamically adjusts its focus to emphasize more relevant feature maps while suppressing noise. This aligns with the principle of selectively amplifying signals that represent the angles of arrival.

The final stage of the proposed CNN architecture features fully connected layers that consolidate the spatial features extracted by the preceding convolutional layers. Specifically, the architecture employs three fully connected layers, with the first two layers having dimensions of 2048 and 1024, respectively. To mitigate overfitting, dropout layers with a dropout rate of 0.2 are applied after these layers. The final fully connected layer serves as the network's output, with dimensions determined by the desired number of AoAs to be estimated. For classification purposes, a predefined angular grid is introduced, mapping each grid cell to a unique polar or azimuth angle, facilitating the interpretation of results.

The output layer is structured as a grid where each neuron corresponds to a specific angular position, represented by dimensions $N(T + F)$, where T and F denote the number of neurons allocated for angles θ and ϕ , respectively. Neurons in the grid are classified as either 0 or 1, indicating the absence or presence of an AoA, respectively. This classification process is extended across all neurons, with the total number calibrated to balance the desired accuracy with the classification precision of the neural network.

To address the DoA estimation task as a binary classification problem, a sigmoid activation function is applied to the output layer. This function assigns a probability

score to each neuron, reflecting the likelihood of it representing a specific AoA. Scores closer to 0 denote a low probability, while scores near 1 indicate a high likelihood. The sigmoid activation simplifies the interpretation of the network's predictions, enabling straightforward thresholding to classify neurons accurately. This approach enhances the model's reliability in producing precise and interpretable AoA estimates.

Figure 2 demonstrates the CNN-based DoA estimation model, which consists of three convolutional layers, followed by a spatial attention module and three fully connected layers. The final FC layer outputs the predicted DoA. When adapting to data from a dataset of a different array configuration, the first three convolutional layers are frozen, allowing adaptation by fine-tuning the attention and FC layers.

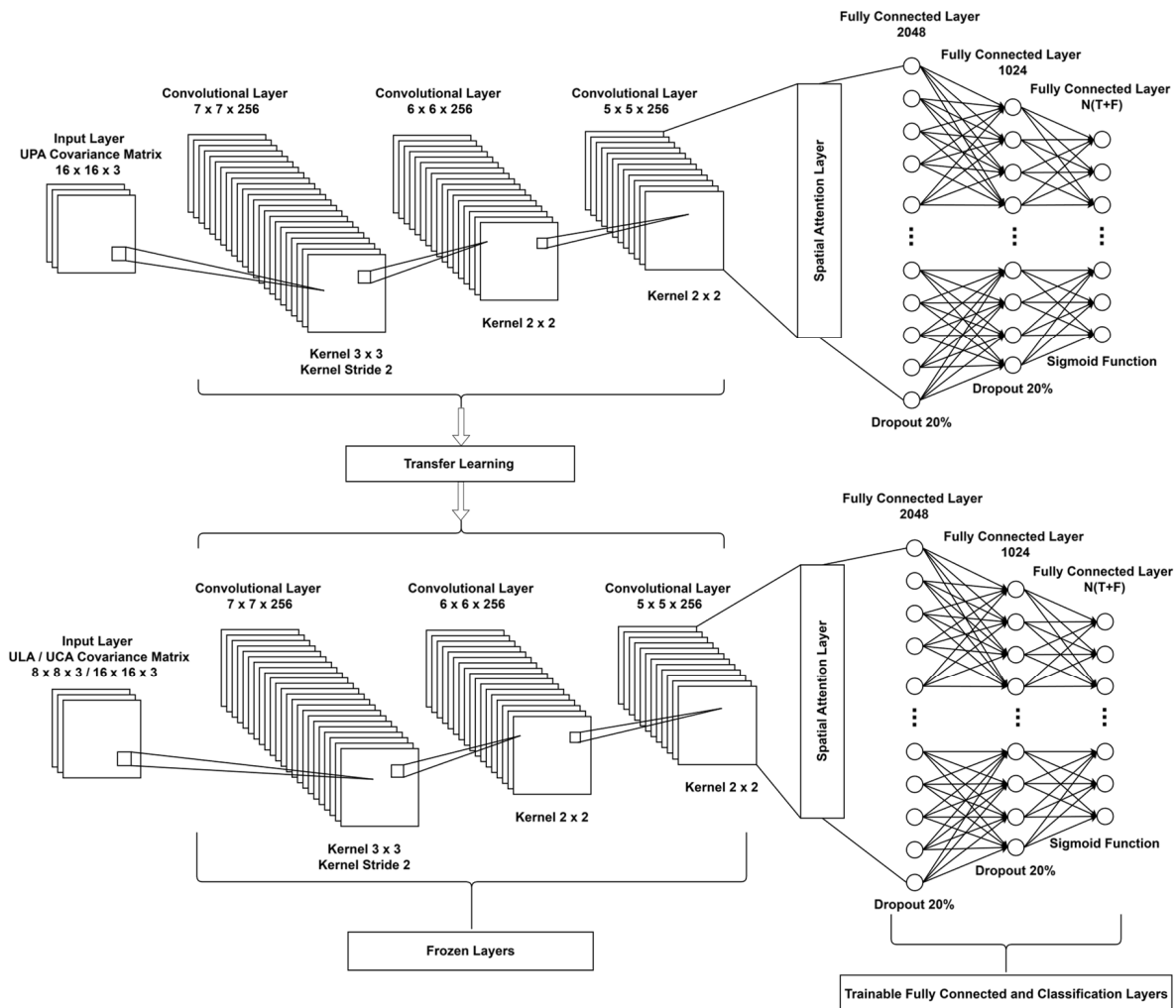


Figure 2. Overview of the proposed CNN-based DoA estimation model architecture, highlighting key elements including spatial attention layers and transfer learning modules.

4.3. Training

To establish a baseline, the model is initially pre-trained using data generated from a UPA, a widely utilized antenna configuration in DoA estimation. This scenario serves as a starting point for evaluating the model's performance and adaptability. In subsequent sections, we will extend this analysis by pre-training the model on alternative configurations, including ULAs and UCAs. During this phase, the model learns core spatial patterns and relationships that are relevant to DoA estimation in the 3D space. Once the model has learned these foundational patterns, transfer learning is applied to adapt it to other array configurations. For deployment on different array configurations, the model undergoes selective fine-tuning: initial layers, which capture fundamental spatial features, are kept

frozen, while higher layers specific to the new configuration are fine-tuned. This process allows for rapid adaptation with minimal additional data and computational resources.

The figures below illustrate the convergence and performance stability of the proposed model over training epochs, highlighting both training and validation accuracy and training and validation loss. In Figure 3, the training and validation accuracy curves demonstrate steady improvement and convergence as the number of epochs increases. Both curves start at lower accuracy levels and show rapid improvement in the initial epochs, indicating the model's capacity to quickly learn essential features. As training progresses, the accuracy levels approach convergence, suggesting that the model has effectively learned the task with minimal overfitting, as seen by the alignment of the validation curve with training accuracy toward the final epochs.

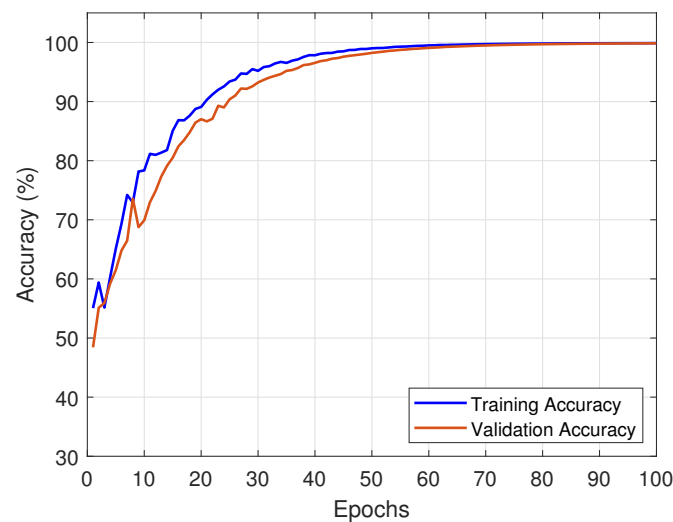


Figure 3. Training and validation accuracy over epochs.

Figure 4 provides a complementary view of model stability through the evolution of training and validation loss over epochs. The loss curves start at higher values and decrease significantly in the initial stages, reflecting effective model optimization. The validation loss closely follows the training loss, further supporting that the model is generalizing well to unseen data. As epochs progress, the loss values stabilize at a low level, confirming the robustness of the proposed model and the successful minimization of errors.

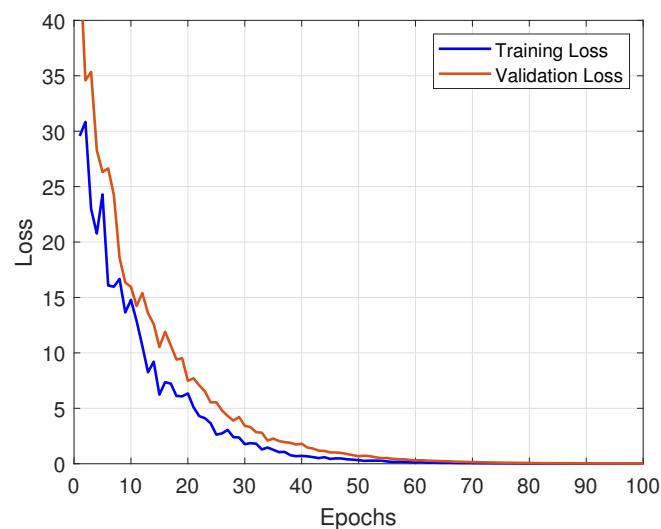


Figure 4. Training and validation loss over epochs.

5. Results and Discussion

In this section, a detailed analysis of the results obtained from the proposed model is presented. Statistical analysis relies on a 1° interval discretized output grid, introducing a baseline MAE of around 0.23° . This error comes from the finite resolution of the grid, reflecting a compromise between computational efficiency and output precision. Despite being an inherent artifact of the discretization approach, maintaining errors near this threshold is key to showcasing the model's learning capacity and its proficiency in reducing estimation errors. This analysis underscores the balance between grid granularity and the model's accuracy in angle-of-arrival prediction. The evaluation spans multiple aspects, including the effectiveness of transfer learning, accuracy across configurations, the impact of spatial attention, robustness under noise, computational efficiency, and comparison to state-of-the-art models.

5.1. Effectiveness of Transfer Learning

Firstly, the accuracy is evaluated across all antenna array configurations, capturing the adaptability to differing spatial layouts. The proposed model demonstrates high accuracy in estimating DoAs. In Table 1, the results show the mean MAE and RMSE achieved by the proposed model for two incoming signals across configurations and random-noise environments with SNR ranging from -20 to 20 dB. The transfer learning framework achieves remarkably low training times while maintaining strong generalization across configurations. This highlights the model's ability to adapt to new setups with minimal performance degradation, enabling swift deployment in diverse scenarios. Its ability to fine-tune effectively for various array configurations with minimal retraining effort underscores the practical value of transfer learning for real-world applications.

Table 1. Effectiveness of transfer learning in cross-configuration generalization. Mean error (MAE, RMSE, and MAPE) for the proposed model across different antenna configurations and random-noise environments.

Antenna Array Configuration	Pre-Training	Fine-Tuning Training Time (min)	MAE ($^\circ$)	RMSE ($^\circ$)	MAPE (%)
ULA	UPA	5:21	0.25	0.28	0.45
UCA	UPA	6:17	0.23	0.32	0.42
UPA	ULA	6:43	0.26	0.29	0.54
UCA	ULA	4:52	0.28	0.30	0.68
UPA	UCA	6:19	0.29	0.34	0.67
ULA	UCA	4:45	0.30	0.36	0.83

To further substantiate the effectiveness of transfer learning in our framework, it is important to align our practical results with the theoretical understanding of transfer learning, which emphasizes the leveraging of knowledge from a source domain to improve learning in a target domain. The transfer learning paradigm is beneficial in settings where labeled data are scarce, and computational efficiency is essential, as it enables faster convergence with fewer labeled samples [35]. In the context of our work, the transfer learning approach allows the model to retain learned representations from a pre-trained configuration and then fine-tune it for specific configurations with minimal effort. As shown in Table 1, this not only maintains strong performance in terms of MAE, RMSE, and MAPE, but also drastically reduces training time. This aligns with the theoretical premise that transfer learning minimizes retraining efforts and accelerates the model's ability to adapt to new, unseen data, thereby enabling swift deployment across varying spatial configurations. By effectively fine-tuning the model for new antenna layouts, we maintain

high generalization capabilities across different environments, further underscoring the practical and theoretical advantages of transfer learning in DoA estimation.

5.2. Impact of Spatial Attention

The integration of spatial attention has been instrumental in refining the model's focus on regions of the input that hold high signal relevance, especially in high-noise environments. To quantify this impact, we assess the MAE reduction attributable to spatial attention at various SNR levels, demonstrating a marked improvement over the baseline CNN model without attention mechanisms. Table 2 details the performance increase achieved by spatial attention in diverse signal environments for two incoming signals, highlighting its role in noise reduction and enhanced feature extraction.

Table 2. Impact of spatial attention across different SNR levels.

SNR (dB)	MAE (°) Without Attention	MAE (°) with Attention
−20	0.60	0.76
−15	0.48	0.52
−10	0.40	0.36
−5	0.36	0.32
0	0.33	0.30
5	0.27	0.26
10	0.24	0.24
15	0.24	0.23
20	0.23	0.23

The observed trend indicates that the performance improvement attributed to spatial attention is more significant at lower SNR levels and gradually diminishes as SNR increases. This behavior can be explained by considering the role of spatial attention in mitigating noise interference. At lower SNR values, the received signal is heavily contaminated by noise, making it more challenging for the neural network to extract meaningful spatial features. The introduction of spatial attention enhances feature selection by emphasizing the most relevant signal components while suppressing noisy or less informative regions. This targeted feature enhancement results in a notable reduction in MAE at low SNR.

However, as SNR increases, the input signal quality naturally improves, and the neural network can already extract spatial features with high fidelity even without attention. In such high-SNR scenarios, the additional benefit of attention mechanisms diminishes because the baseline model without attention is already approaching the optimal estimation performance. This is particularly evident for SNR values above 10 dB, where the MAE difference between the models with and without attention is minimal. This trend aligns with theoretical expectations, where the effectiveness of noise-mitigation techniques becomes less significant in high-SNR conditions.

5.3. Impact of Dataset Size on Accuracy and Computational Complexity

The dataset size plays a critical role in balancing model accuracy and computational efficiency. Larger datasets provide more training diversity, improving generalization, but they also increase computational costs, memory requirements, and training time. To analyze this tradeoff, we evaluate the model's performance across different dataset sizes ranging from 10^4 to 10^8 records.

Table 3 presents the MAE, training time, and memory usage for different dataset sizes.

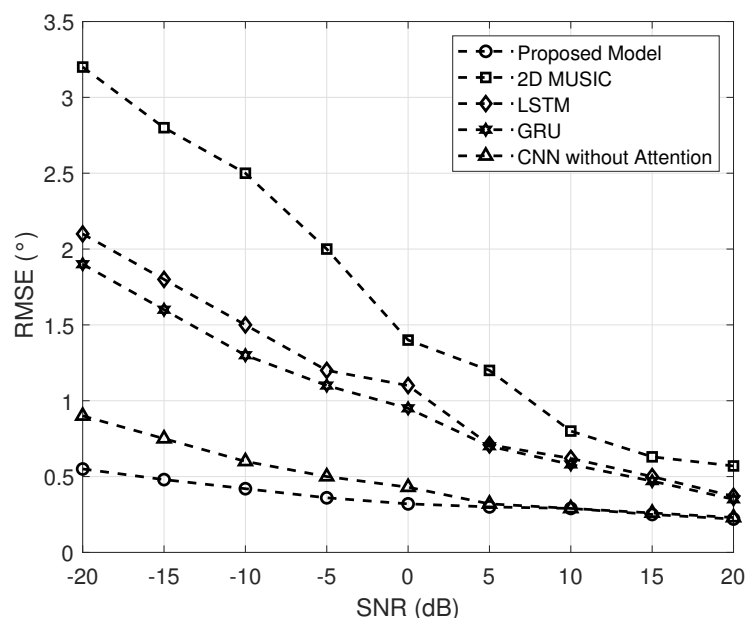
Table 3. Impact of dataset size on accuracy and computational complexity, on estimating two targets.

Dataset Size	MAE (°)	Training Time (h)	Memory Usage (GB)
10^4	0.45	1.2	3.5
10^5	0.34	3.8	8.2
10^6	0.23	9.5	42.7
10^7	0.23	14.3	196.4
10^8	0.23	72.8	985.2

The dataset size plays a crucial role in model performance, but accuracy gains become limited due to the resolution of the DoA estimation grid. Increasing the dataset size significantly improves accuracy up to 10^6 records. However, beyond this point, the MAE remains stable at 0.23° , despite substantial increases in training time and memory usage. This plateau occurs because the estimation accuracy is inherently constrained by the predefined grid resolution, limiting further improvements regardless of additional training data. Thus, while larger datasets improve robustness, there is no practical benefit in exceeding 10^6 records. This dataset size strikes the optimal balance between accuracy, computational efficiency, and real-time feasibility.

5.4. Sensitivity to Noise and Interference

The robustness of the model to noise was evaluated by evaluating the performance across a range of SNR levels, from -20 dB to 20 dB. The results in Figure 5 demonstrate the accuracy of the model and the low error rates even at low SNR, which emphasizes its suitability for environments with considerable noise and interference. The spatial attention mechanism contributed significantly to the model's resilience, allowing it to retain high accuracy in challenging noise conditions.

**Figure 5.** Comparison against conventional and state-of-the-art models.

5.5. Complexity

The computational complexity of the proposed model was analyzed by parameter count and memory requirements. The addition of spatial attention layers did not significantly increase overall complexity, as shown in Table 4. This efficiency, combined with the benefits of transfer learning, allows the model to maintain high performance without excessive resource demands, making it suitable for embedded systems.

Table 4. Model complexity in terms of total number of parameters and memory usage.

Model Component	Number of Parameters (Million)	Memory (MB)
Baseline CNN	3.3	11.7
CNN with Attention	3.5	12

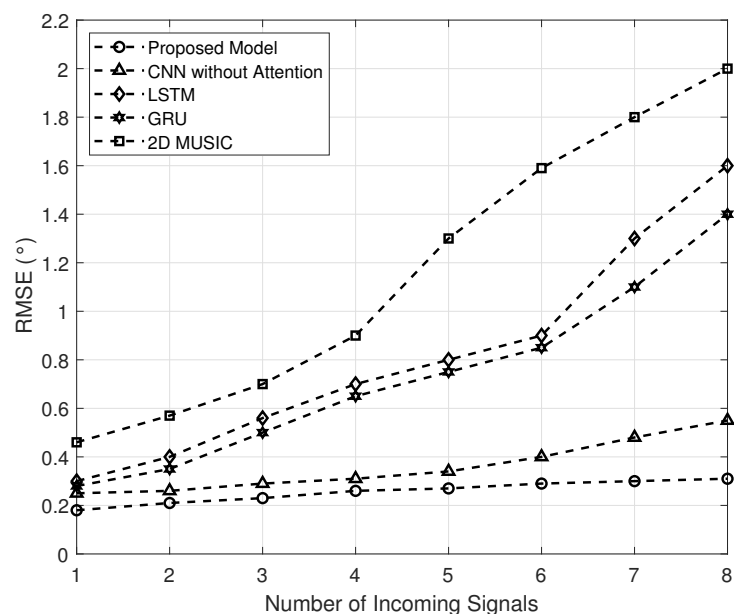
To assess the real-time feasibility of the model, we measured inference times for both batch and single-instance scenarios. The results, as presented in Table 5, demonstrate that our model maintains lower inference times compared to the baseline model, largely due to the spatial attention layers that optimize feature extraction.

Table 5. Inference efficiency in batch and single-instance scenarios.

Scenario	Baseline Inference Time (ms)	Proposed Model Inference Time (ms)
Batch (32 instances)	175	151
Single Instance	90	82

5.6. Performance Comparison

Figure 6 shows a comparison of RMSE between the proposed model and conventional and state-of-the-art models on different numbers of incoming signals. The plot clearly indicates that the CNN model achieves superior accuracy in all tested scenarios, with the greatest advantage in multiple incoming signals. The superior performance can be attributed to the model's ability to capture complex, non-linear relationships in the data.

**Figure 6.** Comparison of RMSE between the proposed model, conventional, and state-of-the-art models against multiple incoming signals.

The experimental results provide compelling evidence of the adaptability and precision of the proposed model in various antenna configurations and challenging signal conditions. One notable observation is the effectiveness of the model in leveraging transfer learning to reduce training times while maintaining high accuracy, as shown in Table 1. This feature is particularly relevant for scenarios where frequent reconfiguration of antenna arrays is necessary, such as in dynamic communication systems or mobile networks. Moreover, the model's fine-tuning capability underscores its potential for broader applicability, demonstrating that it is not constrained by the specificities of any one array configuration.

Spatial attention mechanisms emerge as a pivotal factor in the model's resilience to noise and interference. Table 2 highlights consistent MAE reductions across all SNR levels, emphasizing the mechanism's ability to identify and prioritize regions of the input covariance matrix that are most relevant for DoA estimation. This finding not only validates the inclusion of spatial attention, but also suggests a path for further refinement; potentially by exploring hybrid attention mechanisms that could combine spatial and channel-wise focus. The marked improvement under low-SNR conditions also underscores the model's potential in environments where traditional techniques struggle, such as in highly cluttered or urban communication scenarios.

From a computational point of view, the analysis shows that the added complexity of spatial attention remains minimal, as seen in Table 4. However, this small trade-off is offset by significant gains in accuracy and inference efficiency. Although inference times are slightly reduced, this improvement hints at optimized processing during the feature extraction phase, enabled by the spatial attention layers. These findings open up opportunities for the deployment of the model in real-time applications, including edge computing scenarios where computational resources are limited. Additionally, the model's superior performance in handling multiple incoming signals (Figure 6) sets it apart from existing methods, making it a strong candidate for next-generation communication systems that require robust multi-signal processing capabilities.

Unlike [27,28], which focus primarily on transformer-based models, our framework incorporates spatial attention layers optimized for DoA estimation. While transformers excel in capturing long-range dependencies, their computational overhead is significantly higher. Our approach leverages spatial attention within a CNN backbone to efficiently capture spatial correlations while maintaining computational feasibility. Compared to [29,30], our framework explicitly utilizes a transfer learning mechanism that enables cross-configuration adaptation with minimal retraining. The study in [30] explores transfer learning primarily within a residual attention network, whereas our model applies fine-tuning across multiple antenna configurations, significantly reducing the need for dataset-specific retraining. Moreover, refs. [27–29] rely on gridless or transformer-based architectures, which often require additional computational complexity for robustness against noise. Our model effectively integrates spatial attention and transfer learning, achieving lower error rates, even in low-SNR conditions, as demonstrated in Table 2. This ensures higher accuracy in real-world deployment scenarios with dynamic noise interference. Compared to [30], our model maintains a lower parameter count and faster inference times, as shown in Table 5. This makes it more suitable for real-time applications in embedded systems and edge computing environments.

6. Conclusions

In conclusion, this study presents a robust model architecture that combines deep convolutional neural networks with spatial attention mechanisms and a transfer learning framework to improve DoA estimation across multiple antenna configurations. Key contributions of this work include significant accuracy improvements in DoA estimation in UPAs, ULAs, and UCAs in challenging noise environments. Experimental results demonstrate that our model achieves a lower error in predictions compared to the baseline model, with consistent performance gains across SNRs. Transfer learning is employed to adapt the model to new configurations, with selective fine-tuning minimizing the need for additional data while preserving computational efficiency. By focusing on compact spatial representation and leveraging advanced feature extraction, this approach enables accurate DoA predictions with both stability and responsiveness to complex input variations. Our

findings suggest that it is possible to have DoA estimation with lower prediction errors using the proposed deep learning framework.

Future research could explore several extensions to further refine and enhance the proposed model. The integration of reinforcement learning to enable real-time optimization and continuous model updating in dynamic environments, is promising as it would allow for responsive adaptations to changes in noise and signal characteristics. Finally, the development of more complex spatial attention mechanisms that consider multi-source interference patterns could further improve DoA estimation in densely packed environments, addressing the growing demands of modern wireless communication networks.

Author Contributions: Conceptualization, C.M.M. and Z.D.Z.; methodology, C.M.M.; software, C.M.M. and S.K.G.; validation, C.M.M. and Z.D.Z.; formal analysis, C.M.M. and P.I.L.; investigation, C.M.M. and P.S.; resources, C.M.M. and P.V.; data curation, C.M.M. and P.V.; writing—original draft preparation, C.M.M. and S.K.G.; writing—review and editing, C.M.M. and Z.D.Z.; visualization, C.M.M. and P.I.L.; supervision, Z.D.Z.; project administration, Z.D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the European Union, through the Horizon Europe Marie Skłodowska-Curie Staff Exchanges Programme “6G intelligent connectivity and interaction for users and infrastructures (6G-ICARUS)” under Grant 101131342.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Pantelis Velanas was employed by the company Acceligen Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
DoA	Direction of arrival
DL	Deep learning
ML	Machine learning
MIMO	Multiple-input multiple-output
MAE	Mean absolute error
NN	Neural network
LSTM	Long short term memory
RMSE	Root mean squared error
SNR	Signal to noise ratio
SGD	Stochastic gradient descent
UCA	Uniform circular array
ULA	Uniform linear array
UPA	Uniform planar array

References

1. Schoeder, P.; Janoudi, V.; Meinecke, B.; Werbunat, D.; Waldschmidt, C. Flexible Direction-of-Arrival Simulation for Automotive Radar Target Simulators. *IEEE J. Microw.* **2021**, *1*, 930–940. [[CrossRef](#)]
2. Lee, H.; Ahn, J.; Kim, Y.; Chung, J. Direction-of-Arrival Estimation of Far-Field Sources Under Near-Field Interferences in Passive Sonar Array. *IEEE Access* **2021**, *9*, 28413–28420. [[CrossRef](#)]
3. Van Veen, B.D.; Buckley, K.M. Beamforming: A Versatile Approach to Spatial Filtering. *IEEE ASSP Mag.* **1969**, *5*, 4–24. [[CrossRef](#)] [[PubMed](#)]

4. Capon, J. High-Resolution Frequency-Wavenumber Spectrum Analysis. *Proc. IEEE* **1969**, *57*, 1408–1418. [[CrossRef](#)]
5. Schmidt, R.O. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [[CrossRef](#)]
6. Roy, R.; Kailath, T. ESPRIT—Estimation of Signal Parameters via Rotational Invariance Techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995. [[CrossRef](#)]
7. Wax, M.; Kailath, T. Spatial Spectral Analysis Using the Eigenstructure Method. *IEEE Trans. Acoust. Speech Signal Process.* **1985**, *32*, 817–827. [[CrossRef](#)]
8. Hawes, M.; Mihaylova, L.; Septier, F.; Godsill, S. Bayesian compressive sensing approaches for direction of arrival estimation with mutual coupling effects. *IEEE Trans. Antennas Propag.* **2017**, *65*, 1357–1368. [[CrossRef](#)]
9. Friedlander, B.; Weiss, A.J. Performance of direction-finding systems with sensor gain and phase uncertainties. *Circuits Syst. Signal Process.* **1993**, *12*, 3–35. [[CrossRef](#)]
10. Swindlehurst, A.L.; Stoica, P. Maximum likelihood methods in radar array signal processing. *Proc. IEEE* **1998**, *86*, 421–441. [[CrossRef](#)]
11. Ottersten, B.; Kailath, T. A performance analysis of subspace-based methods in the presence of model errors. I. The MUSIC algorithm. *IEEE Trans. Signal Process.* **2002**, *40*, 1758–1774.
12. Mallioras, I.; Yioultsis, T.V.; Kantartzis, N.V.; Lazaridis, P.I.; Zaharis, Z.D. Enhancing Adaptive Beamforming in 3-D Space Through Self-Improving Neural Network Techniques. *IEEE Open J. Commun. Soc.* **2024**, *5*, 1340–1357. [[CrossRef](#)]
13. Rekkas, V.P.; Sotiroudis, S.; Sarigiannidis, P.; Wan, S.; Karagiannidis, G.K.; Goudos, S.K. Machine Learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends. *Electronics* **2021**, *10*, 2786. [[CrossRef](#)]
14. Iliadis, L.A.; Zaharis, Z.D.; Sotiroudis, S.; Sarigiannidis, P.; Karagiannidis, G.K.; Goudos, S.K. The road to 6G: A comprehensive survey of deep learning applications in cell-free massive MIMO communications systems. *J. Wireless Com. Netw.* **2022**, *2022*, 68. [[CrossRef](#)]
15. Vijayamohanan, J.; Gupta, A.; Noakoasteen, O.; Goudos, S.K.; Christodoulou, C.G. Source Detection with Multi-Label Classification. *IEEE Open J. Signal Process.* **2023**, *4*, 336–345. [[CrossRef](#)]
16. Raiguru, P.; Kumar Rout, S.; Sahani, M.; Kishore Mishra, R. Machine Learning-Aided Sparse Direction of Arrival Estimation. *IEEE Sens. J.* **2024**, *24*, 38125–38134. [[CrossRef](#)]
17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
18. Xu, Y.; Zhang, L.; Liu, F.; Zhang, Y. Feed-Forward Neural Network Approach for Real-Time DOA Estimation in High-Mobility Scenarios. *Electronics* **2021**, *10*, 2531.
19. Hu, S.; Zeng, C.; Liu, M.; Tao, H.; Zhao, S.; Liu, Y. Robust DOA Estimation Using Deep Complex-Valued Convolutional Networks with Sparse Prior. In Proceedings of the 2023 6th International Conference on Information Communication and Signal Processing (ICICSP), Xi'an, China, 23–25 September 2023; pp. 234–239.
20. Lu, Y.; Guan, H.; Yang, K.; Peng, T.; Wen, C.; Li, X. Improving the Accuracy of Direction of Arrival Estimation with Multiple Signal Inputs Using Deep Learning. *Sensors* **2024**, *24*, 2971. [[CrossRef](#)]
21. Liu, Z.; Gao, W.; Zhang, T.; Zhang, J. Direction of Arrival Estimation Using Convolutional Neural Networks with Spatial Spectrum Data. *IEEE Trans. Veh. Technol.* **2021**, *70*, 6639–6652.
22. Mylonakis, C.M.; Velanas, P.; Lazaridis, P.I.; Sarigiannidis, P.; Goudos, S.K.; Zaharis, Z.D. 3D Direction of Arrival Estimation: An Innovative Deep Neural Network Approach. In Proceedings of the 2024 13th International Conference on Modern Circuits and Systems Technologies (MoCAST), Sofia, Bulgaria, 26–28 June 2024.
23. Wang, S.; Yu, F.; Zhang, X.; Liang, Q. Recurrent Neural Network-Based DOA Estimation for Dynamic Channels in Vehicular Communications. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 2110–2120.
24. Ye, K.; Zhou, L.; Hong, S.; Zhang, X.; Sun, H. Research on 2-D Direction of Arrival (DOA) Estimation for an L-Shaped Array. *Remote Sens.* **2024**, *16*, 4787. [[CrossRef](#)]
25. He, J.; Fakhreddine, A.; Vanwynsberghe, C.; Wymeersch, H.; Alexandropoulos, G.C. 3D Localization with a Single Partially-Connected Receiving RIS: Positioning Error Analysis and Algorithmic Design. *IEEE Trans. Veh. Technol.* **2023**, *72*, 13190–13202. [[CrossRef](#)]
26. Takemoto, R.; Cha, J.; Jeong, I.; Ahn, C.-J. Effectiveness and Characteristics of Virtual Antennas in the Multiple Signal Classification Algorithm. *Electronics* **2025**, *14*, 73. [[CrossRef](#)]
27. Vougioukas, G.; Bletsas, A. DoA Estimation with a Single Antenna and a Few Low-Cost Backscattering Tags. *IEEE Trans. Commun.* **2022**, *70*, 6849–6860. [[CrossRef](#)]
28. Xu, S.; Brighente, A.; Conti, M.; Chen, B.; Wang, S. Evaporative Angle: A Generative Approach to Mitigate Jamming Attacks in DOA Estimation. *IEEE Wirel. Commun. Lett.* **2024**, *13*, 2922–2926. [[CrossRef](#)]
29. Zhang, Q.; Tang, Y.; Li, J.; Han, J. Hybrid CNN-RNN Model for Robust Direction of Arrival Estimation in LOS and NLOS Environments. *Sensors* **2022**, *22*, 1034.

30. Lee, H.; Kim, S.; Park, Y. Transformer-Based Deep Learning for Direction of Arrival Estimation in Complex Signal Environments. *IEEE Access* **2023**, *11*, 10567–10576.
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017.
32. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. *arXiv* **2021**, arXiv:2112.05561.
33. Mallioras, I.; Yioultis, T.V.; Kantartzis, N.V.; Lazaridis, P.I.; Zaharis, Z.D. A Novel Neural Network Approach to Proactive 3-D Beamforming. *IEEE Trans. Cogn. Commun. Netw.* **2024**, early access.
34. Wu, X.; Wang, J.; Yang, X.; Tian, F. A Gridless DOA Estimation Method Based on Residual Attention Network and Transfer Learning. *IEEE Trans. Veh. Technol.* **2024**, *73*, 9103–9108. [[CrossRef](#)]
35. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
36. Cao, H.; Xiong, C.; Gong, L.; Jiang, T. Deep Transfer Learning for Underwater Direction of Arrival Using One Vector Sensor. *J. Acoust. Soc. Am.* **2021**, *149*, 1699–1711. [[CrossRef](#)] [[PubMed](#)]
37. Cao, B.; Xiong, C.; Gong, L.; Jiang, T. Two-Dimensional Direction of Arrival Estimation Based on Nested Circular Array. In Proceedings of the 2022 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), Denver, CO, USA, 10–15 July 2022; pp. 124–125.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.