

## Research Article

# Deep Learning Approach in DOA Estimation: A Systematic Literature Review

Shengguo Ge <sup>1</sup>, Kuo Li <sup>2</sup>, and Siti Nurulain Binti Mohd Rum<sup>1</sup>

<sup>1</sup>Faculty of Computer Science & Information Technology, Universiti Putra Malaysia, 43400 Serdang, Malaysia

<sup>2</sup>Construction Machinery Division, CMIK (Beijing) Vehicle Testing and Engineering Research Institute Co., Beijing 100000, China

Correspondence should be addressed to Kuo Li; 1536174196@qq.com

Received 19 July 2021; Revised 16 August 2021; Accepted 8 September 2021; Published 18 September 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Shengguo Ge et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In array signal processing, the direction of arrival (DOA) of the signal source has drawn broad research interests with its wide applications in fields such as sonar, radar, communications, medical detection, and electronic countermeasures. In recent years, the application of deep learning (DL) to DOA estimation has achieved great success. This study provides a systematic review of research on DOA estimation using deep neural network methods. We manually selected twenty-five papers related to this research from five prominent databases (SpringerLink, IEEE Xplore, ScienceDirect, Scopus, and Google Scholar) for exploration. Six questions describing the overall trend of DOA estimation using deep learning are put forward. Then, we answered these questions by reviewing the literature. This review is helpful for researchers in this field because it provides more specific and comprehensive information needed for future research. Specifically, we first analyzed the background of the selected papers, including the type of publication, the number of citations, and the country of origin. Then, the DL technology used in DOA estimation is systematically analyzed, including the purpose of using DL in DOA estimation, various DL models (convolutional neural network, deep neural network, and combination network), and various DOA estimation schemes. Finally, various evaluation criteria (root-mean-squared error, accuracy, and mean absolute error) are used to evaluate the DL technology in DOA estimation, and various factors (signal-to-noise ratio, number of snapshots, number of antennas, and number of signal sources) affecting DOA estimation are analyzed. Based on our findings, we believe that deep learning can perform DOA estimation well, and there is still room for improvement in deep learning technology. In this study, the factors affecting DOA estimation can be used as the direction for researchers to conduct in-depth research.

## 1. Introduction

Early DOA estimation has its origins in the conventional beamforming (CBF) [1], which directly corresponds to the traditional Fourier spectrum estimation method from the time domain signal processing method to the spatial domain signal processing, such that the array angle resolution is restricted by the Rayleigh limit constraint. The Rayleigh limit means that the two signals can be distinguished only when the angular separation between the two far-field signal sources is greater than the antenna beamwidth. The Capon method [2] can minimize the output energy in the interference direction while keeping the output energy in the desired direction constant. This method does not require the

number of sources in advance and is robust, but its resolution is not high enough. Eigen subspace methods can break the Rayleigh limit. The Pisarenko method is a harmonic analysis method [3]. It obtains the signal subspace and noise subspace by performing eigenvalue decomposition or singular value decomposition on the array covariance matrix and uses the orthogonality between each other. The eigenvector corresponding to the smallest eigenvalue is taken as the noise subspace, and a high-precision DOA estimation of the target is obtained with a small computational cost. However, this algorithm has limitations because it is only suitable for the number of array elements that exceeds the number of signal sources by one. The DOA estimation is obtained with super-resolution capabilities, the

most promising of which are the multiple signal classification (MUSIC) method [4] based on noise subspace and the estimation of signal parameters by the rotation invariance technology (ESPRIT) [5].

Traditional DOA estimation methods are generally based on a physical array structure, so they have certain limitations. For virtual antenna array technology, the signal received by the actual array antenna can be used to construct the signal of the virtual array element [6]. It increases the degree of freedom of the array and expands the array aperture, which allows it to handle more signal sources and improve the performance of DOA estimation. This technology mainly includes the interpolation transformation method and the fourth-order statistics method. The interpolation transformation method divides the transformation area according to the approximate direction of the received signal from the array antenna. Then it obtains the transformation matrix according to the virtual transformation array flow pattern vector and the actual array flow pattern vector and calculates the virtual transformation covariance matrix, ultimately achieving the virtual array transformation [7]. The fourth-order statistics method treats the fourth-order statistics of the data received by the array as a cross-correlation between the specific actual array elements and the corresponding virtual array elements. Then, it replaces the second-order covariance matrix with the constructed fourth-order covariance matrix [8]. These methods further enhance the performance of DOA estimation because they break through the limitations of physical arrays on DOA estimation. In addition, there is a DOA estimation algorithm based on the compressed sensing theory, which converts the DOA estimation problem into a row sparse matrix reconstruction problem and reduces the amount of calculation by singular value decomposition.

In practical applications, the number of signal sources is usually unknown. However, except for a few methods (such as the Capon method), most DOA estimation methods require prior knowledge of the number of signal sources. If the estimated number of signal sources does not match the actual number, the orthogonality of the signal subspace and noise subspace will be affected. This severely affects the accuracy of DOA estimation. In response to this limitation, many scholars have proposed classical methods for estimating the number of signal sources. Akaike [9] proposed a method based on the Akaike information criterion (AIC). Wax and Ziskind [10] proposed a method based on the minimum description length (MDL). However, the AIC and MDL criteria are only applicable to Gaussian white noise. In a practical environment, colored noise is more common than Gaussian white noise. Therefore, these two methods are not applicable in the condition of colored noise. By contrast, the RAIC [11] and RMDL [12] estimation methods using diagonal loading technology in the information theory criterion can effectively suppress the divergence of noise eigenvalues to smooth color noise, with good estimation performance under any form of array. The Gerschgorin disk estimation (GDE) method proposed by Wu et al. [13] can estimate the number of signal sources for an unknown array signal against white noise and colored noise background.

The GDE method does not obtain the signals from all sensors, so when the number of sources is large, the probability of correct detection will decrease rapidly due to insufficient degrees of freedom.

Although researchers have proposed a number of DOA estimation methods, a method that can accurately estimate DOA has yet to be found. With the development of artificial intelligence (AI), estimating DOA with AI has the potential to yield a remarkable result. Compared with traditional algorithms, the DOA estimation algorithm based on deep learning greatly improves the estimation performance and generalization, but there is still room for improvement in performance under high noise or strong reverberation conditions. In this paper, we present a systematic literature review (SLR) on the approach of deep learning in DOA estimation computation. This paper is organized as follows. Section 2 is the introduction of deep learning. Section 3 presents the methodology used to perform the systematic literature review. Section 4 analyzes the current research status of DOA estimation. Section 5 examines the different architectures of DL used in DOA estimation. Section 6 compares the performance of various methods of DOA estimation using deep learning. Finally, the paper is concluded in Section 7.

## 2. Deep Learning

Machine learning is a process by which machines use artificial neural networks (ANNs) to compute large amounts of Internet-based data to autonomously simulate the process of human learning and ultimately make smart decisions. ANN is an adaptive nonlinear dynamic network system composed of a large number of neurons connected to each other. It is a simulation and approximation of biological neural networks. As shown in Figure 1, the MP model [14] is the first mathematical model of neurons. Later, Rosenblatt [15] used the linear optimization method to simulate the nervous system of human learning and proposed a single-layer perceptron model. However, their model cannot deal with the linear inseparability problem. To make up for this shortfall, Rumelhart et al. [16] proposed the backpropagation (BP) neural network, which is a multilayer feedforward neural network trained according to the error backpropagation algorithm. However, when the number of network layers increases, the BP neural network was found to have problems such as local optimization, overfitting, and gradient diffusion, which restrict the use of the BP neural network. In addition, some other excellent shallow neural networks have emerged, such as support vector machines (SVM) [17] and Gaussian mixture models (GMM) [18]. But they fail to address problems of machine learning such as overfitting [19], the curse of dimensionality [20], and theoretical guarantees [21]. To solve these problems, a new branch of machine learning was produced: deep learning. Because such learning does not require artificial design or feature extraction and can learn from data autonomously, it is also called unsupervised learning.

Deep learning originated from machine learning and statistical mechanics. Based on the discrete Hopfield

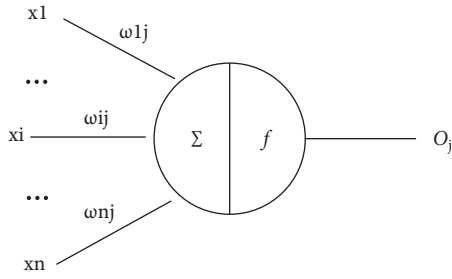


FIGURE 1: MP model.

network [22], Ackley et al. [23] used Boltzmann distribution to propose the Boltzmann machine (BM), which introduces statistical probability into the neuron state changes. The equilibrium state of the network follows the Boltzmann distribution, and the optimal solution is sought through the simulated annealing algorithm. In the following year, Smolensky [24] proposed a restricted Boltzmann machine (RBM) by defining BM as a two-layer network, namely, the visible unit layer and the hidden unit layer. Besides, it is stipulated that the neurons in different layers are independent of each other while those in the same layer are connected to each other [25], as shown in Figure 2. Hinton and Salakhutdinov [26] proposed to connect RBMs in series to form a deep belief network (DBN) consisting of multiple stacked RBMs and a backpropagation (BP) network. The training process is divided into two steps: pretraining and fine-tuning. In the pretraining process, the output of the RBM of the previous layer is used as the input for the next layer, and the unsupervised greedy method is used to train each RBM from bottom to top with the weight being updated during training. The weight obtained after the pretraining is used as the initial weight of the DBN, and then the entire network is adjusted by the error between the top-level output and the expected output. The number of hidden layers of RBM is increased to obtain a deep Boltzmann machine (DBM), which uses the contrast divergence CD algorithm [27] to greatly improve the training speed of the DBM. The network structures of DBN and DBM are shown in Figure 3 [26].

Convolutional neural networks (CNNs) are currently more popular than other artificial neural networks. The basic structure of the CNN consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer [28]. The convolutional layer is composed of many feature maps, with each map consisting of many neurons. Each neuron is connected to a part of the functional graph of the previous layer by a convolution kernel that serves as a weight matrix [29]. Then the weighted sum of the local area is passed to a nonlinear function to obtain the output value of each neuron in the convolutional layer. The weights of the CNN are shared between the same input feature map and the output feature map. The convolutional layer of the CNN extracts different features of the input through a convolution operation, with the first layer extracting low-level features and the higher-level one extracting more advanced features. The convolutional layer

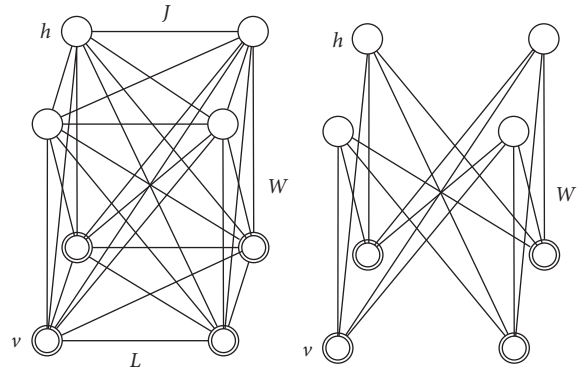


FIGURE 2: General Boltzmann machine and restricted Boltzmann machine.

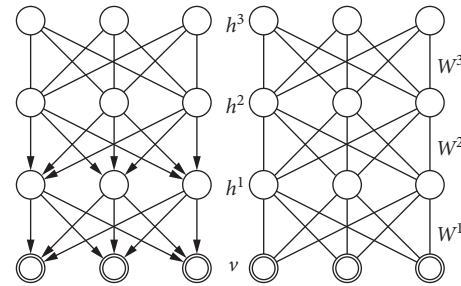


FIGURE 3: Deep belief network and deep Boltzmann machine.

is followed by the pooling layer, which is also composed of many feature maps. Each of the maps uniquely corresponds to each feature map of the convolutional layer, so the number of feature maps is constant. The pooling layer, which plays the role of secondary extraction of features, obtains spatially invariant features by reducing the resolution of the feature map [30]. Each neuron in this layer performs pooling operations on the local receptive field. Commonly used pooling methods include maximum pooling, which is to take the point with the largest value in the local receptive field, and average pooling, which is to take the average of the values in the locally accepted domain [31]. Among the commonly used pooling methods, overlapping pooling methods can be used even when different neurons in the same feature map of the pooling layer do not overlap with the local receptive field of the previous layer. For this pooling method, there is an overlapping area between adjacent pooling windows [32]. The pooling layer reduces the number of neurons through the pooling operation and reduces the amount of calculation of the network model. In a CNN, after several convolutional layers and several pooling layers, one or more fully connected layers are connected. In the fully connected layer, which is also called a softmax layer, each neuron is fully connected to all neurons in the previous layer, and the activation function of the neurons in it is generally the ReLU function [33]. This layer can integrate category-discriminatory information in the convolutional layer or the pooling layer [34]. Then, its output value is passed to an output layer for classification using softmax regression [35].

Recurrent neural networks (RNNs) can be used to process and predict sequence data. It overcomes many limitations of traditional machine learning methods on input and output data and can be well applied when there is certain time dependence in the data. Therefore, RNN is an important model for deep learning. The long short-time memory (LSTM) network proposed by Hochreiter and Schmidhuber [36] is now the most effective sequence model in practical applications. LSTM is improved on the basis of RNN. LSTM uses three gate structures to control the state and output at different times, namely, input gate, output gates, and forget gate. The LSTM architecture mentioned by Graves and Schmidhuber [37] is shown in Figure 4. LSTM mitigates gradient disappearance by combining short-term memory with long-term one through a gate structure. The gate structure can be regarded as a fully connected layer that stores and updates information, and its activation function is the sigmoid function. The Sigmoid function will output a value between 0 and 1 to indicate the number of messages that can pass through the door at the current moment. Zero means that no information can be passed, and 1 means that all information can be passed. The forget gate is a key component of the LSTM unit, which controls which information should be retained and which should be forgotten, thus avoiding the gradient disappearance and the explosion caused by gradient backpropagation over time. The input gate controls the amount of information that flows into the memory unit from the current input data, and the output gate determines the amount of information that the memory unit outputs in the current state.

### 3. Review Methodology

In this research work, twenty-five papers related to the topic have been selected. Five prominent databases including SpringerLink, IEEE Xplore, ScienceDirect, Scopus, and Google Scholar have been explored. Six questions describing the overall trend of DOA estimation using deep learning are put forward in the research questions section. Based on the guidelines provided by Kitchenham and Charters [38] (as shown in Figure 5), three main steps have been undertaken to carry out this SLR work, namely, planning, conducting, and reporting, and each step consists of several activities.

**3.1. Research Questions.** The proposed system evaluation aims to research how DOA estimates can benefit from the application of DL technology. The research questions are listed below:

- RQ1: What types of publication distributions available from the databases over the last seven years related to the topic area?
- RQ2: Why is DL technology applied to DOA estimation?
- RQ3: Which DL techniques are applied to DOA estimation?
- RQ4: What are the key aspects of DOA estimation?

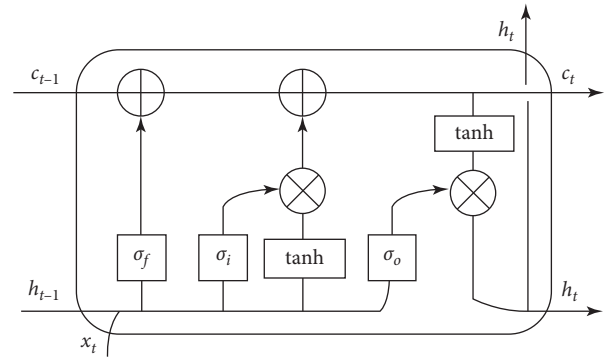


FIGURE 4: LSTM unit structure.

RQ5: What evaluation criteria are used for DOA estimation, and how do they perform?

RQ6: What factors affect DOA estimation, and how do they affect DOA estimation?

**3.2. Study Identification and Selection.** The searched phrases are divided into two groups: DOA estimation and deep learning. The string is defined as any term related to signal processing (e.g., “DOA estimation,” and “source number estimation,” “source number enumeration,” and “direction of arrival,” combined with the function OR) with any term related to deep learning (e.g., “artificial neural network,” “human convolutional neural network,” “deep learning,” “CNN,” “DNN,” and “RNN”). The search platform chosen were EZAccess Portal (Malaysia Putra University Library Database) and Google Scholar. The former portal contains many well-known databases, namely, SpringerLink, IEEE Xplore, ScienceDirect, and Scopus. The latter portal has a wide range of academic literature, making it easier to search.

Using the above string combination in the database, a total of 2,499 papers are returned. Table 1 shows the distribution of papers in each database. The duplicated papers are deleted, and a total of 2,444 papers are excluded using the exclusion criteria (not written in English, repetitions, books, inaccessible papers, works not in the shipping area, and papers less relevant to the research direction of the review). During the extraction process, the remaining 55 papers are analyzed. Given the great number of papers, we analyzed the abstracts of these papers, and the 25 most relevant papers were selected for review.

### 4. Analysis of Current Research

This section aims to introduce the basic information of related papers and provide the answer for RQ1 based on the following findings:

- (1) Figure 6 shows the number of related articles from 2015 to March 2021. As can be seen, since 2017, the use of deep learning in DOA estimation studies has become increasingly popular.
- (2) The number of journal publications from each country is presented in Figure 7. Among them, papers from China account for the vast majority,

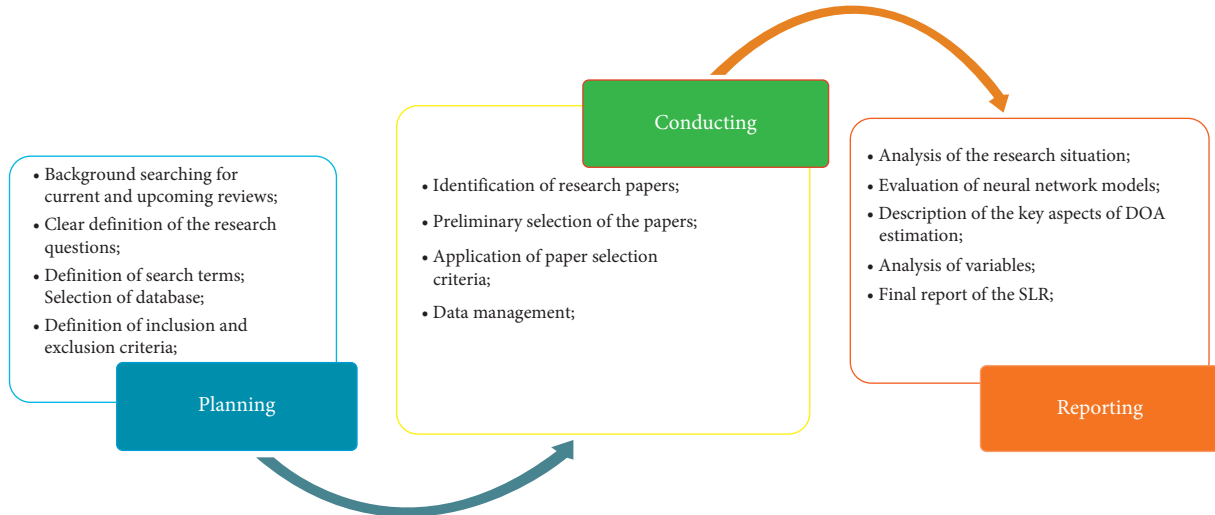


FIGURE 5: A systematic literature review process [38].

TABLE 1: Number of papers retrieved from each database.

Database	Number of articles found
SpringerLink	129
IEEE Xplore	67
ScienceDirect	133
Scopus	891
Google Scholar	1,280

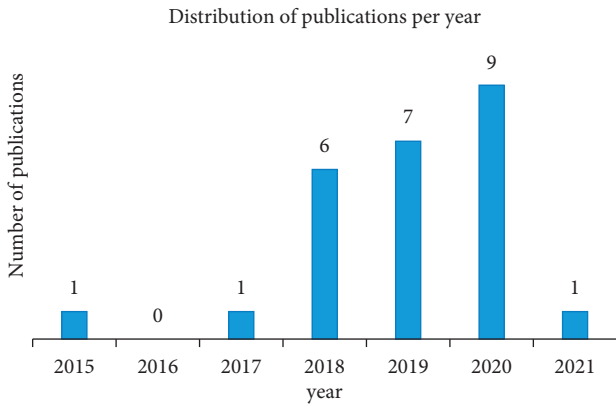


FIGURE 6: Publications per year.

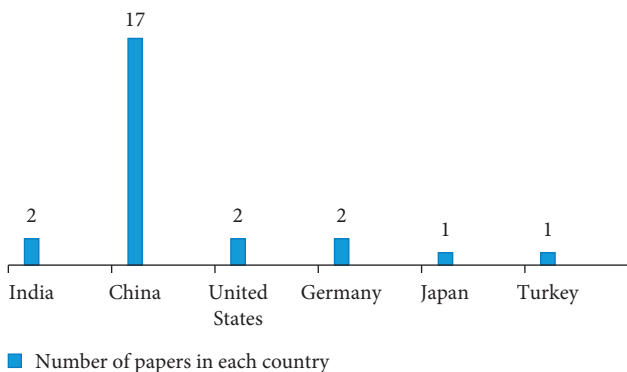


FIGURE 7: Publications per country.

with a total of 17 papers, showing that China attaches the most importance to this research direction. This is followed by two papers each from India, the USA, and Germany, and one each from Japan and Turkey.

- (3) Table 2 shows our statistics on the number of citations and publication types for the 25 related papers. As can be seen from the table, 7 papers were published in conferences, and the other 18 papers were published in journals. We included 7 conference papers because there were not enough journal papers to select as the application of deep learning in DOA estimation is not yet widespread.
- (4) Citation count is a reference indicator for published papers. The more the citation count, the more valuable the work is supposed to be. Table 3 provides statistics on the number of times relevant papers have been cited. It can be seen from the table that there are 14 papers with a citation count less than 5, 2 papers with citations greater than or equal to 5 and less than 10, and 9 papers with a citation count greater than or equal to 10. It can be seen that nearly half of the papers have citation counts greater than 5. In addition, it should be noted that some papers have a zero citation count due to their short publication period. Among them, the work done by Huang et al. [57] have been cited 321 times; the work done by Chakrabarty and Habets [61] have been cited 114 times; the work done by Liu et al. [58] have been cited 83 times; and the work done by Chakrabarty and Habets [54] have been cited 70 times.

## 5. Deep Learning in DOA Estimation

Research on related papers shows that different DL models frameworks and algorithms are applied to DOA estimation. This section answers the research questions RQ2, RQ3, and RQ4. To this end, the advantages of DL applied to DOA estimation compared to traditional DOA estimation

TABLE 2: Publication type of selected papers (citation data via Google Scholar, checked on March 1, 2021).

Source	Publication Amount of citations	Type of publication	
		Journal	Conference
Liu et al. [39]	0	✓	
Varanasi et al. [40]	1	✓	
Xiao et al. [41]	3	✓	
Shi et al. [42]	0	✓	
Chen et al. [43]	0	✓	
Elbir [44]	12	✓	
Yao et al. [45]	2	✓	
Cong et al. [46]	0	✓	
Xiang et al. [47]	1	✓	
Zhu et al. [48]	0	✓	
Rogers et al. [49]	0		✓
Yang et al. [50]	6	✓	
Xiang et al. [51]	6	✓	
Goodman et al. [52]	0		✓
Fu et al. [53]	2	✓	
Chakrabarty and Habets [54]	70	✓	
Wajid et al. [55]	3		✓
Pan et al. [56]	3	✓	
Huang et al. [57]	321	✓	
Li et al. [31]	15		✓
Liu et al. [58]	83	✓	
Kase et al. [59]	13		✓
Wang et al. [60]	12	✓	
Chakrabarty and Habets [61]	114		✓
Zheng et al. [62]	10		✓

TABLE 3: Number of citations for the considered papers over 2015–2021.

Citation count	<5	≥5 and <10	≥10
Number of papers	14	2	9

methods are emphasized. Then, we introduced various DL techniques used in DOA estimation. Finally, we summarized various scenarios in DOA estimation.

*5.1. Purpose of Using DL.* To answer RQ2, we analyzed the disadvantages of the traditional DOA estimation method and the advantages of the DL method. The traditional DOA estimation algorithm is generally limited by many factors. For example, the received coherent signal will cause the signal subspace and the noise subspace to permeate each other, which reduces the accuracy of many classical subspace-like DOA estimation algorithms. In addition, the resolution of the DOA estimation algorithm will be limited by the physical aperture of the array, and the maximum number of resolvable targets will be limited by the number of elements. Besides, in practical applications, the number of signal sources is often unknown, and most DOA estimation methods need to know the number of signal sources in advance. Otherwise, the estimation of the signal subspace

and the noise subspace will be inaccurate, which affects the orthogonality of the signal subspace and the noise subspace and ultimately DOA estimation accuracy.

Compared with traditional methods, the DL method converts DOA estimation into pattern recognition. The DOA estimation is carried out by extracting the features of the signal data, which overcomes the disadvantages of the traditional DOA estimation algorithm and improves the accuracy of the DOA estimation.

*5.2. Deep Learning Techniques.* To answer RQ3, the DL techniques used were reviewed from several papers. Table 4 is the summarization of the array model, DL model, activation function, number of network layers, and description of the network structure used in the considered papers.

The information given in Figures 8–10 provides the answer to RQ3. Figure 8 summarizes the array model used in DOA estimation. It can be seen from the figure that ULA is used the most (81%), followed by UCA (14%), and SMA (5%) is also used in speech DOA estimation. Figure 9 is the distribution of DL techniques applied in the research work of the selected reviewed papers. The most used DL model is CNN (40%), followed by DNN (36%). Then, some combination networks (12%) are used, such as CNN-RNN, CNN-LSTM, and DNN-SVM. There are also some rarely used network models (12%), such as DFN, RNN, and SVM. Figure 10 summarizes the activation functions used in the DL model for DOA estimation. It can be seen from the figure that the most used activation functions are ReLU (57%), followed by Sigmoid (19%), Tanh (10%), and others (14%).

In summary, ULA is the most frequently used array model by researchers; CNN and DNN are the most frequently used DL models; and ReLU is the most commonly used activation function.

### 5.3. Different Aspects of DOA Estimation

*5.3.1. DOA Estimation in Signal Processing.* DOA estimation is an important task in signal processing, and it has a wide range of applications in fields such as radar, sonar, and so on. Xiang et al. [51] analyzed the unknown multipath signal and concluded that the unknown multipath signal severely distorts the phase characteristic distribution of the desired signal. The designed supervised deep neural network is used for phase enhancement, thereby effectively reducing the phase distortion. The verification of real data shows that this method effectively improves the DOA estimation. Goodman et al. [52] evaluated two new techniques for estimating the direction of arrival of RF sources: constrained integer optimization and deep learning. Research has found that deep learning is more robust to significant calibration errors. To adapt to the DOA estimation in the urban environment, Shi et al. [42] proposed a complex-valued convolutional neural network (CCNN). Experiments show that CCNN has a faster convergence rate than CNN and a higher DOA estimation accuracy. Xiang et al. [47] proposed a supervised CNN phase enhancement model, which can reduce phase distortion by enhancing the phase characteristics. Their

TABLE 4: General description of the systematized papers.

References	Array model	Technique	Activation function	Convolutional layers	Description of architecture
Liu et al. [39]	ULA	CNN	Sigmoid	9	The depth of the network and the number of convolution and pooling layers will increase as the input increases.
Varanasi et al. [40]	SMA	CNN	ReLU	27	It consists of two different CNNs: one CNN is used to estimate the elevation angle and the other CNN is used to estimate the azimuth angle.
Xiao et al. [41]	ULA	CNN	Tanh	8	The network structure is similar to the deep residual network.
Shi et al. [42]	ULA	CNN	ReLU	—	The existing CNN is converted into a complex-valued convolutional neural network to gain more degrees of freedom.
Chen et al. [43]	ULA	DNN	ReLU	4	DNN consists of an input layer, a hidden layer, and an output layer. The number of hidden layers is 2.
Elbir [44]	ULA	CNN	ReLU	17	CNN has more network layers than signals, and each layer corresponds to an angular spectrum.
Yao et al. [45]	ULA	CNN and RNN	Sigmoid	—	The network structure consists of a convolutional layer, a BiLSTM layer, and a fully connected feedforward layer.
Cong et al. [46]	ULA	DNN	Tanh and Sigmoid	—	The framework includes an autoencoder, a feedforward network, a network parameter database, and a collection of a series of directed acyclic graph networks.
Xiang et al. [47]	ULA	CNN	ReLU	—	CNN consists of two convolutional layers, a nonlinear layer and a linear layer, and then the maximum pooling layer.
Zhu et al. [48]	UCA	CNN	ReLU	5	Each convolutional layer has 128 feature maps; the fully connected layer has 512 neurons; and the output layer uses three neurons.
Rogers et al. [49]	ULA	DNN	ReLU	15	DNN has an input layer, four sets of fully connected layers, and then a batch normalization layer. Before the last fully connected layer, there is a dropout layer, followed by a softmax and classifier layer.
Yang et al. [50]	ULA	DNN	ReLU	10	DNN is composed of a regression network and classification network.
Xiang et al. [51]	ULA	DNN	ReLU	4	Each layer contains 1,000 neurons.
Goodman et al. [52]	UCA	DFN	$\exp^{-x^2}$	6	NFN includes two components: the internal feedback interaction component and an external feedback interaction component.
Fu et al. [53]	ULA	CNN	—	—	Integrate deep learning and convolutional nonnegative matrix factorization for DOA estimation.
Chakrabarty and Habets [54]	ULA	CNN	—	7	CNN consists of an input layer, three hidden layers, two fully connected layers, and an output layer.
Wajid et al. [55])	ULA	RNN	—	—	RNN is based on bidirectional long short-term memory (BiLSTM) networks.
Pan et al. [56]	UCA	SVM	—	—	The MATLAB toolbox “libsvm” is used to determine the best parameters.
Huang et al. [57]	—	DNN	Sigmoid and ReLU	8	The hidden layer of DNN can be divided into coding layer, dropout layer, noise layer, and decoding layer according to its function.
Li et al. [31]	—	CNN and LSTM	ReLU	7	The model consists of three convolutional layers, one LSTM layer, and a full connection layer.
Liu et al. [58]	ULA	DNN	Linear	—	The network consists of two parts: one is a multitasking autoencoder and the other is a fully connected multilayer neural network.
Kase et al. [59]	ULA	DNN	Ramp	—	DNN is stacked by several single-layer neural networks.
Wang et al. [60]	—	DNN and SVM	—	7	DNN contains three hidden layers, which are cascaded with SVM for multisource DOA estimation.
Chakrabarty and Habets [61]	ULA	CNN	ReLU	7	The number of convolutional layers of CNN is one, less than the number of microphones in the array.
Zheng et al. [62]	—	DNN	—	5	DNN has 3 hidden layers, and each hidden layer has 512 neurons.

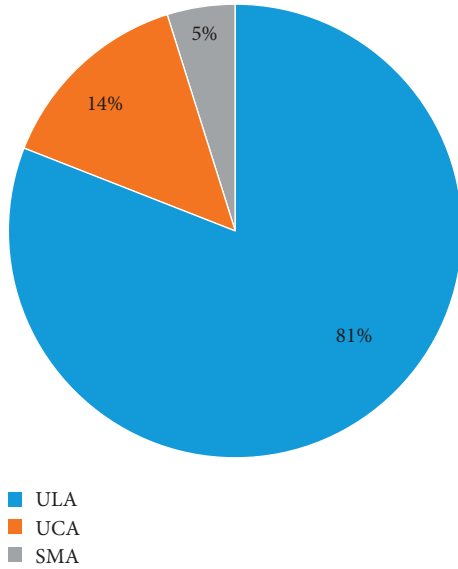


FIGURE 8: Distribution of array models of related papers.

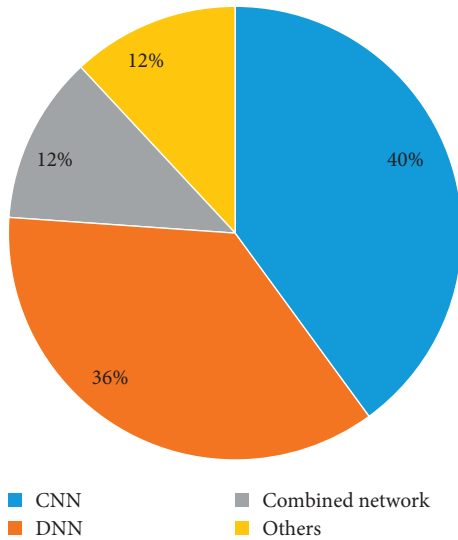


FIGURE 9: Distribution of DL technique of related papers.

model can effectively improve the DOA estimation accuracy of the very high frequency (VHF) radar system under coherent multipath signals. Zhu et al. [48] proposed to train five CNNs under different conditions, with each CNN learning the direction image converted from the real and imaginary parts of the spatial covariance matrix. Then, the prediction results of each CNN are calculated as an average value to obtain the final DOA estimation results. Simulation shows that this method can perform effective two-dimensional (2D) DOA estimation. Cong et al. [46] proposed a DNN-based DOA estimation framework that includes an autoencoder, a feedforward network, a network parameter database, and a collection of a series of directed acyclic graph networks (DAGN). Among them, the autoencoder is equivalent to the noise filter, and each subnet of DAGN is composed of a convolutional neural network (CNN) and

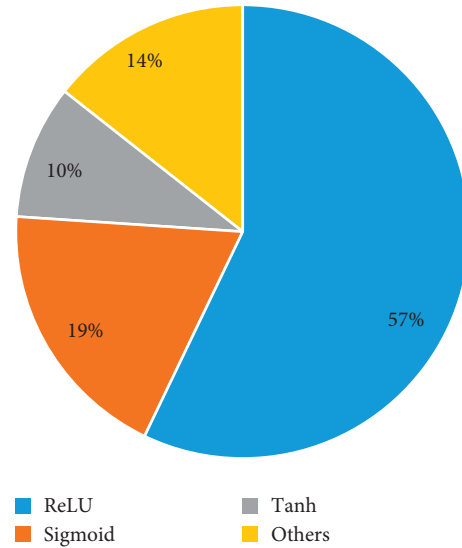


FIGURE 10: Distribution of DL activation functions of related papers.

two bidirectional long short-term memory (BiLSTM) networks. The simulation shows that the DOA estimation performance of this network is better than the traditional subspace algorithm. Yao et al. [45] proposed a DOA estimation model based on a recurrent neural network. With the help of Toeplitz matrix reconstruction, the model can estimate DOA for signals with unknown signal sources. However, this model does not perform well in an environment with a low signal-to-noise ratio and color noise. Elbir [44] designed multiple CNNs, and such that each CNN is dedicated to an angular spectrum to learn the multiple signal classification (MUSIC) spectra of the corresponding angle subregion. This method reduces the amount of calculation and improves the accuracy of DOA estimation. Kase et al. [59] designed a stacked DNN with multiple single-layer neural networks. The lower triangular part of the correlation matrix of the received signal vector is used as input to train the DNN. The simulation results show that the DNN designed in a specific scenario has good DOA estimation performance. Liu et al. [39] designed multiple CNNs based on the number of array elements and used covariance matrices containing real and imaginary numbers for training. After a large amount of data learning, this method can effectively identify the direction of underwater acoustic signals. Liu et al. [58] proposed a DNN framework, which consists of a multitask autoencoder and a series of parallel multilayer classifiers. The encoder and the classifier are trained on different data sets. The function of the autoencoder is to decompose the input into multiple components in different spatial subregions. Simulation shows that this method can be well applied to array defects, but in practical applications, it faces a significant challenge as it requires a large amount of labelled data for training. Chen et al. [43] proposed a DNN framework for DOA estimation of radio waves. The network is divided into a detection network and a DOA estimation network. For the detection network, the search area of the antenna array is divided into several



sectors, with each of them corresponding to a DOA estimation network. This is to detect the signal radiated by each sector. According to the detection result, one or more DOA estimation networks can be activated for DOA estimation. Simulation shows that compared with the traditional method, this method streamlines the calculation, improves the estimation accuracy, and has excellent generalization ability. Huang et al. [57] proposed a novel DNN framework for super-resolution DOA estimation and channel estimation through offline learning and online learning. Among them, offline learning is to use simulated data for training under different channel conditions, and online learning is to obtain the corresponding output data based on the current input data. Experiments have confirmed that methods based on deep learning can achieve better DOA estimation than traditional methods. Xiao et al. [41] proposed a DeepFPC network structure similar to the deep residual network. DeepFPC has high sparse signal recovery performance and good DOA estimation performance under low SNR.

**5.3.2. Speech DOA Estimation.** Speech DOA estimation can be applied to distant automatic speech recognition. Varanasi et al. [40] proposed a network architecture applied to DOA estimation by discussing the azimuth angle and source height existing in the amplitude and phase characteristics of spherical harmonics. They also expanded the DOA estimation method into a dense DOA search grid. Training and testing were performed using data sets of simulated and real environments, respectively, and performance evaluation showed that DOA estimation was improved even in noisy and reverberant environments. Fu et al. [53] proposed a new blind DOA estimation method that uses the 2D convolution nonnegative matrix factorization method to generate a new array signal to estimate the azimuth angle of the reverberation signal. Wajid et al. [55] proposed to use the recurrent neural network (RNN) model to learn some similar features used in DAS beamforming. The results show that the DOA estimation result based on RNN is better than DAS beamforming. Li et al. [63] developed a supervised learning algorithm combining CNN and long short-term memory (LSTM) network for DOA estimation. Retraining the model with new data makes the method robust to noise and reverberation and can quickly adapt to new microphone arrays. Chakrabarty and Habets [54] proposed to use the supervised learning method of CNN to estimate the DOA of multiple speakers. This method formulates the multispeaker DOA estimation as a multiclass multilabel classification problem. Among them, the characteristics of each input element are regarded as a separate binary classification problem. This method can accurately locate the speaker in a dynamic acoustic scene. Zheng et al. [62] proposed to use different values of SNR and noise to train DNNs, which achieved higher DOA accuracy at low SNR and improved the intersensor data ratio (ISDR) performance of a single acoustic vector sensor (AVS) in a noisy environment. Wang et al. [60] proposed to use acoustic vector sensors (AVSs) to estimate the DOA of multiple voice sources through clustering of data ratios between sensors. This method designs a

connection between DNN and SVM. Using a soft mask learner, the time-frequency points (TD-TFP) dominated by the target speech can be extracted under different noisy and reverberation conditions, thus improving the estimation performance. Chakrabarty and Habets [61] proposed to use CNN as a classifier for wideband DOA estimation. Their method uses the synthesized noise signal to train the CNN. It directly feeds the phase component of the short-time Fourier transform coefficients of the received microphone signal into the CNN and performs wideband DOA estimation. Experimental evaluation shows that the method has good robustness to noise and small disturbances.

**5.3.3. Signal Source Number Estimation.** Most of the existing DOA estimation algorithms require prior knowledge of the signal source number, so the estimation of the number of signal sources is the primary task of DOA estimation. If the estimated quantity is different from the actual one, the DOA estimation will be affected. Pan et al. [56] proposed a source number enumeration model (M-UCA) of UCA with  $M$  antennas. This model can estimate the number of signal sources at most  $M-1$  by extracting features from the instantaneous phase of the array signal and then using SVM as a classifier to classify signals with different numbers of signal sources. Yang et al. [50] proposed to use a regression network (ERNet) and a classification network (ECNet) for source number detection. Then the signal's covariance matrix is taken as input and the number of signal sources as data labels for training. This data-driven method can automatically learn the threshold used to separate signal and noise characteristic values and does not require a Gaussian hypothesis for derivation like traditional methods. Simulation experiments have verified the effectiveness of this method. Rogers et al. [49] designed a 15-layer deep learning network with parameter correction linear units, which uses eigenvalues and spatial smoothing covariance matrix entries as inputs to estimate the number of sources of narrowband signals. Although this literature survey is preliminary, it plausibly suggests that deep learning can have better source number estimation capabilities.

## 6. Performance Metrics and Explanatory Variables

This section aims to answer the research questions RQ5 and RQ6. For this purpose, the performance metrics of related papers are counted, and the factors affecting DOA estimates are analyzed.

**6.1. Performance Metrics.** Evaluating the performance of the DL model is important for verifying the quality of the DOA estimation algorithm. To answer RQ5, we summarized the performance of the DL model in the considered papers. As shown in Table 5, some performance criteria are used to evaluate the performance of the DL model.

Figure 11 helps answer RQ5 by counting the distributions of different evaluation criteria used by researchers. The most commonly used evaluation criteria are root-mean-squared

TABLE 5: Parameter comparison and performance evaluation during DOA estimation.

Reference	SNR (dB)	Number of snapshots	Number of sensors	Number of sources	Evaluation index	Evaluation value
Liu et al. [39]	10	—	10	1	Accuracy	98%
Varanasi et al. [40]	1	—	32	3	GE	0.14
Xiao et al. [41]	15	10	40	3	MAE	0.6
Shi et al. [42]	10	8	8	1	Accuracy	74.8%
Chen et al. [43]	2	400	10	2	RMSE	0.2
Elbir [44]	25	100	16	6	RMSE	0.02
Yao et al. [45]	10	100	11	3	MAE	0.95
Cong et al. [46]	25	400	5	2	RMSE	0
Xiang et al. [47]	14	42	21	2	RMSE	0.1
Zhu et al. [48]	10	1,024	9	1	RMSE	0.59
Rogers et al. [49]	3	20	11	5	Accuracy	0.77
Yang et al. [50]	5	100	10	—	Accuracy	100%
Xiang et al. [51]	3	48	24	1	RMSE	0.02
Goodman et al. [52]	5	100	6	1	ME	0.028
Fu et al. [53]	10	20	3	1	RMSE	0
Chakrabarty and Habets [54]	5	—	4	2	MAE	0.6
Wajid et al. [55]	14	—	4	1	AA	1
Pan et al. [56]	20	500	4	3	Accuracy	98%
Huang et al. [57]	15	64	128	32	ME	0.001
Li et al. [31]	6	—	6	—	Accuracy	74.8%
Liu et al. [58]	10	400	10	2	RMSE	0.2
Kase et al. [59]	30	100	5	2	RMSE	2.6
Wang et al. [60]	15	—	—	2	RMSE	2.6
Chakrabarty and Habets [61]	30	—	4	2	MAE	1.5
Zheng et al. [62]	20	—	4	1	RMSE	0

RMSE: root-mean-squared error; MAE: mean absolute error; MSE: mean squared error; GE: gross error; ME: mean error; A: accuracy; AA: average accuracy.

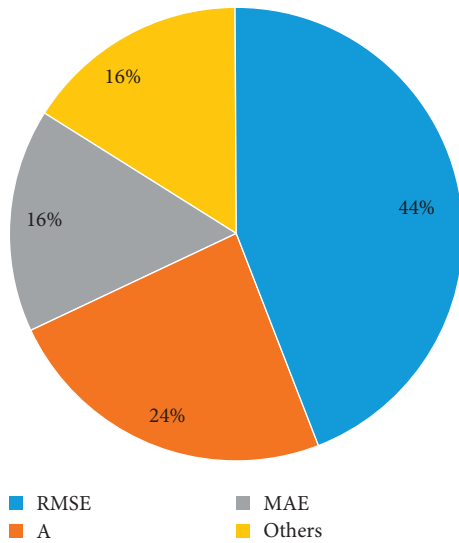


FIGURE 11: Distribution of evaluation criteria in the considered papers.

error (RMSE), accuracy (A), and mean absolute error (MAE), accounting for 44%, 24%, and 16%, respectively. In addition, there are some other evaluation criteria that are less used, such as mean squared error (MSE), gross error (GE), mean error (ME), and average accuracy (AA). The following is a brief description of the frequently used criteria.

The root-mean-squared error is used to measure the deviation between the observed value and the true value. The

average absolute error can avoid the mutual cancellation of errors, so it can accurately reflect the size of the actual forecast error. We assume: (i)  $y = \text{true value}$ ; (ii)  $h(x) = \text{observed value}$ ; and (iii)  $m = \text{number of observations}$ . Then the expressions of RMSE and MAE are as follows:

$$\text{RMSE}(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2},$$

$$\text{RMSE}(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)})^2}, \quad (1)$$

$$\text{MAE}(X, h) = \frac{1}{m} \sum_{i=1}^m |h(x^{(i)}) - y^{(i)}|.$$

Accuracy refers to the ratio of the number of samples correctly predicted to the total number of predicted samples. It does not consider whether the predicted samples are positive or negative. For the two-category confusion matrix, the implications of the four parameters are as follows: (i) true positives (TP): both the predicted and actual samples are positive; (ii) false positives (FP): positive predicted samples and negative actual samples; (iii) true negatives (TN): both predicted and the actual samples are negative; (iv) false negatives (FN): negative

TABLE 6: The estimated performance of each method under different SNRs.

Xiang et al. [51]	SNR	-5 dB	0 dB	5 dB
	RMSE	0.08	0.03	0.02
Fu et al. [53]	SNR	-5 dB	0 dB	5 dB
	RMSE	3.3	0.6	0.3
Chen et al. [43]	SNR	-10 dB	0 dB	10 dB
	RMSE	1.3	0.3	0.28

predicted samples and positive actual samples. The expression can be expressed as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (2)$$

The gross error is an error other than a random error or systematic error. The mean error is the average of all errors in a group. The average accuracy refers to the average of the accuracy rates of all categories.

By comparing each method, Fu et al. [53] had the best performance when using RMSE as the evaluation method, and RMSE is 0. When using A as the evaluation method, Yang et al. [50] had the best performance, and A is 100%. When using MAE as the evaluation method, Chakrabarty and Habets [54] had the best performance, and MAE is 0.6. However, the performance of different methods is subject to the influence of various factors, which lead to performance variations. The various factors affecting DOA estimates will be described in the next section.

**6.2. Factors Affecting the DOA Estimation.** The result of DOA estimation is related to the incident signal source and the environment in which it is applied. This section aims to answer RQ6, so we first count the various factors that affect the DOA estimation and then analyze the results of each factor's influence on the DOA estimation.

The DL method of DOA estimation is affected not only by the DL models but also by the data set. The data set is determined by the characteristics of the incident signal. Table 6 counts the various factors affecting DOA estimation in the papers reviewed, including signal-to-noise ratio (SNR), as well as the number of snapshots, antennas, and signal sources. The following is an analysis of how each factor affects DOA estimation.

**6.2.1. The Impact of SNR on DOA Estimation.** SNR directly affects the performance of the super-resolution DOA estimation algorithm. As shown in Table 6, increasing the SNR from -5 dB to 5 dB in [51] reduces its RMSE from 0.08 to 0.02. Increasing the SNR from -5 dB to 5 dB in [53] reduces its RMSE from 3.3 to 0.3, while the RMSE drops from 1.3 to 0.28 when increasing the SNR from -10 dB to 10 dB in [43]. From the above relationship between SNR and RMSE, it can be seen that as the signal's SNR increases, the DOA estimation performance of the algorithm also improves. Therefore, improving the DOA estimation performance of the algorithm under low SNR conditions is the primary task of the high-resolution DOA estimation algorithm.

TABLE 7: The estimated performance of each method under different number of snapshots.

Yao et al. [45]	Snapshots	20	60	100
	MAE	1.6	1.0	0.1
Cong et al. [46]	Snapshots	50	200	400
	RMSE	0.4	0.21	0.1
Yang et al. [50]	Snapshots	10	100	1,000
	Accuracy	91%	99%	100%

TABLE 8: The estimated performance of different methods under different  $M/L$ .

Xiao et al. [41]	$M/L$	1/20	3/20	5/20
	MAE	0.35	0.50	1.02
Xiang et al. [51]	$M/L$	1/20	—	1/10
	Accuracy	100%	—	97%
Elbir [44]	$M/L$	1/8	—	3/8
	RMSE	0.01	—	0.018

Note:  $M$  is the number of signal sources and  $L$  is the number of array antennas.

**6.2.2. The Impact of the Number of Snapshots on DOA Estimation.** The number of snapshots is defined differently: in the time domain, it is the number of sampling points; in the frequency domain, it is the number of time subsegments of the discrete Fourier transform. As shown in Table 7, if the number of snapshots in [45] increases from 20 to 100, the MAE drops from 1.6 to 0.1. Increasing the number of snapshots from 50 to 400 in [46] reduces the RMSE from 0.4 to 0.1. Increasing the number of snapshots from 10 to 1,000 in [50] increases the accuracy from 91% to 100%. The above relationship between the number of snapshots and the evaluation criteria shows that the DOA estimation performance improves as the number of snapshots increases.

**6.2.3. The Impact of the Number of Antennas and the Number of Signal Sources on DOA Estimation.** The number of antennas in the array and the number of sources of the incident signal also affect the DOA estimation. Table 8 shows the relationship between the ratio of the number of signal sources to the number of antennas and the evaluation criteria. If the ratio in [41] increases from 1/20 to 5/20, the MAE increases from 0.35 to 1.02. In [51], if the ratio increases from 1/20 to 1/10, the accuracy will drop from 100% to 97%; if the ratio in [44] increases from 1/8 to 3/8, the RMSE increases from 0.01 to 0.018. From the above relationship, the DOA estimation performance will improve as the ratio increases. That is to say, the smaller the number of signal sources and the greater the number of antennas, the better the DOA estimation.

## 7. Conclusions

Deep learning (DL) has been successfully applied in many fields due to its powerful capabilities. Therefore, this article presents a systematic literature review of the main papers on DOA estimation using DL technology. This research first conducted a cursory analysis of the 25 selected papers, including the type of publication, the number of citations, and

the country of origin. Then, a systematic analysis of DL techniques used in DOA estimation is then presented, including the purpose of using DL in DOA estimation, various DL models, and various DOA estimation scenarios. Finally, the DL technology in DOA estimation is evaluated, and various factors affecting DOA estimation are analyzed.

*7.1. The Answer of Six Research Questions.* After presenting an overview of the results of this literature review, here, we answer the six research questions given in Section 3.

RQ1: What types of publication distributions available from the databases over the last seven years related to the topic area?

(i) Since 2017, applying deep learning to DOA estimation has become increasingly popular. (ii) Seventeen papers are from China, accounting for a majority of all papers. (iii) Seven papers were published in conferences, and the other eighteen papers were published in journals. (iii) Eleven of the twenty-five papers have an index number of not less than five.

RQ2: Why is DL technology applied to DOA estimation?

The DOA estimation is carried out by extracting the features of the signal data, which overcomes the disadvantages of the traditional DOA estimation algorithm and improves the accuracy of the DOA estimation.

RQ3: Which DL techniques are applied to DOA estimation?

ULA is the most frequently used array model by researchers; CNN and DNN are the most frequently used DL models; and ReLU is the most commonly used activation function.

RQ4: What are the key aspects of DOA estimation?

(i) DOA estimation is an important task in signal processing, and it has a wide range of applications in fields such as radar, sonar, and so on. (ii) Speech DOA estimation can be applied to distant automatic speech recognition. (iii) The estimation of the number of signal sources is the primary task of DOA estimation. If the estimated quantity is different from the actual one, the DOA estimation will be affected.

RQ5: What evaluation criteria are used for DOA estimation, and how do they perform?

The most commonly used evaluation criteria are root-mean-squared error (RMSE), accuracy (A), and mean absolute error (MAE), accounting for 42%, 27%, and 16%, respectively. Fu et al. [53] had the best performance when using RMSE as the evaluation method, and RMSE is 0. When using A as the evaluation method, Yang et al. [50] had the best performance, and A is 100%. When using MAE as the evaluation method, Chakrabarty and Habets [54] had the best performance, and MAE is 0.6.

RQ6: What factors that affect DOA estimation, and how do they affect DOA estimation?

Various factors affecting DOA estimation in the papers are reviewed, including signal-to-noise ratio (SNR), as well as the number of snapshots, antennas, and signal sources. (i) As the signal's SNR increases, the DOA estimation performance of the algorithm also improves. (ii) The DOA estimation performance improves as the number of snapshots increases. (iii) The smaller the number of signal sources and the greater the number of antennas, the better the DOA estimation.

*7.2. Final Remarks.* The application of deep learning technology to DOA estimation has achieved good results. This paper reviews methods for research and achievement. It is hoped that this paper can provide an overview for researchers interested in this field.

## Data Availability

The data generated and analyzed during the current study are available in the public domain.

## Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

- [1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [2] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [3] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical Journal International*, vol. 33, no. 3, pp. 347–366, 1973.
- [4] R. Schmidt and R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [5] M. L. Bencheikh and Y. Wang, "Joint DOD-DOA estimation using combined ESPRIT-MUSIC approach in MIMO radar," *Electronics Letters*, vol. 46, no. 15, pp. 1081–1083, 2010.
- [6] P. Chevalier, L. Albera, A. Ferreol, and P. Comon, "On the virtual array concept for higher order array processing," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1254–1271, 2005.
- [7] S. M. Hosseini and M. A. Sebt, "Array interpolation using covariance matrix completion of minimum-size virtual array," *IEEE Signal Processing Letters*, vol. 24, no. 7, pp. 1063–1067, 2017.
- [8] M. C. Dogan and J. M. Mendel, "Applications of cumulants to array processing. II: non-Gaussian noise suppression," *IEEE Transactions on Signal Processing*, vol. 43, no. 7, pp. 1663–1676, 1995.

- [9] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [10] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 37, no. 8, pp. 1190–1196, 1989.
- [11] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 6, pp. 397–401, 1988.
- [12] N. Ma and J. T. Goh, "Efficient method to determine diagonal loading value," in *Proceedings of the 2003 IEEE International Conference On Acoustics, Speech, and Signal Processing Proceedings (ICASSP'03)*, Hong Kong, China, April 2003.
- [13] H. T. Wu, J. F. Yang, and F. K. Chen, "Source number estimators using transformed Gerschgorin radii," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1325–1333, 1995.
- [14] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [15] F. Rosenblatt, *The Perceptron, A Perceiving and Recognizing Automaton Project Para*, Cornell Aeronautical Laboratory, Buffalo, NY, USA, 1957.
- [16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [18] D. A. Reynolds, *A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification*, ProQuest Dissertations Publishing, Ann Arbor, MI, USA, 1993.
- [19] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM Computing Surveys*, vol. 27, no. 3, pp. 326–327, 1995.
- [20] L. Lerman, R. Poussier, O. Markowitch, and F.-X. Standaert, "Template attacks versus machine learning revisited and the curse of dimensionality in side-channel analysis: extended version," *Journal of Cryptographic Engineering*, vol. 8, no. 4, pp. 301–313, 2018.
- [21] O. Bachem, M. Lucic, and A. Krause, "Practical coresets constructions for machine learning," 2017, <https://arxiv.org/abs/1703.06476>.
- [22] P. Floreen and P. Orponen, *Complexity Issues in Discrete Hopfield Networks*, Aalto University, Espoo, Finland, 1994.
- [23] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Science*, vol. 9, no. 1, pp. 147–169, 1985.
- [24] P. Smolensky, *Information Processing in Dynamical Systems: Foundations of Harmony Theory* University of Colorado Boulder, Boulder, CO, USA, 1986.
- [25] R. Salakhutdinov and G. Hinton, "Deep Boltzmann machines," in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, pp. 448–455, PMLR, Clearwater Beach, FL, USA, 2009.
- [26] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [27] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.
- [28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [30] Y. L. Boureau, N. L. Roux, F. Bach, J. Ponce, and Y. LeCun, "Ask the locals: multi-way local pooling for image recognition," in *Proceedings of the 2011 International Conference On Computer Vision*, pp. 2651–2658, Barcelona, Spain, November 2011.
- [31] C. Li, S. X. Yang, Y. Yang et al., "Hyperspectral remote sensing image classification based on maximum overlap pooling convolutional neural network," *Sensors*, vol. 18, no. 10, p. 3587, 2018.
- [32] Y. Long, F. Zhu, L. Shao, and J. Han, "Face recognition with a small occluded training set using spatial and statistical pooling," *Inform. Sciences*, vol. 430–431, pp. 634–644, 2018.
- [33] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, <https://arxiv.org/abs/1511.08458>.
- [34] T. N. Sainath, A.-R. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *Proceedings of the 2013 IEEE International Conference On Acoustics, Speech And Signal Processing*, pp. 8614–8618, Vancouver, Canada, May 2013.
- [35] L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," *Information Sciences*, vol. 428, pp. 49–61, 2018.
- [36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [37] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks: The Official Journal of the International Neural Network Society*, vol. 18, no. 5–6, pp. 602–610, 2005.
- [38] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering version 2.3," *Engineering*, vol. 45, p. 1051, 2007.
- [39] Y. Liu, H. Chen, and B. Wang, "DOA estimation based on CNN for underwater acoustic array," *Applied Acoustics*, vol. 172, p. 107594, 2021.
- [40] V. Varanasi, H. Gupta, and R. M. Hegde, "A deep learning framework for robust DOA estimation using spherical harmonic decomposition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1248–1259, 2020.
- [41] P. Xiao, B. Liao, and N. Deligiannis, "A deep unfolded network for sparse signal recovery from 1-bit measurements with application to DOA estimation," *Signal Processing*, vol. 176, Article ID 107699, 2020.
- [42] B. Shi, X. Ma, W. Zhang, H. Shao, Q. Shi, and J. Lin, "Complex-valued convolutional neural networks design and its application on UAV DOA estimation in urban environments," *Journal of Communications and Information Networks*, vol. 5, no. 2, pp. 130–137, 2020.
- [43] M. Chen, Y. Gong, and X. Mao, "Deep neural network for estimation of direction of arrival with antenna array," *IEEE Access*, vol. 8, pp. 140688–140698, 2020.
- [44] A. M. Elbir, "Deepmusic: deepMUSIC: multiple signal classification via deep learning," *IEEE Sensors Letters*, vol. 4, no. 4, pp. 1–4, 2020.
- [45] Y. Yao, H. Lei, and W. He, "A-CRNN-based method for coherent DOA estimation with unknown source number," *Sensors*, vol. 20, no. 8, p. 2296, 2020.
- [46] J. Cong, X. Wang, M. Huang, and L. Wan, "Robust DOA estimation method for MIMO radar via deep neural networks," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7498–7507, 2020.

- [47] H. Xiang, B. Chen, T. Yang, and D. Liu, "Phase enhancement model based on supervised convolutional neural network for coherent DOA estimation," *Applied Intelligence*, vol. 50, pp. 2411–2422, 2020.
- [48] W. Zhu, M. Zhang, P. Li, and C. Wu, "Two-dimensional DOA estimation via deep ensemble learning," *IEEE Access*, vol. 8, pp. 124544–124552, 2020.
- [49] J. Rogers, J. E. Ball, and A. C. Gurbuz, "Estimating the number of sources via deep learning," in *Proceedings of the 2019 IEEE Radar Conference (RadarConf)*, pp. 1–5, Boston, MA, USA, April 2019.
- [50] Y. Yang, F. Gao, C. Qian, and G. Liao, "Model-aided deep neural network for source number detection," *IEEE Signal Processing Letters*, vol. 27, pp. 91–95, 2019.
- [51] H. Xiang, B. Chen, M. Yang, T. Yang, and D. Liu, "A novel phase enhancement method for low-angle estimation based on supervised DNN learning," *IEEE Access*, vol. 7, pp. 82329–82336, 2019.
- [52] J. Goodman, D. Salmond, C. Davis, and C. Acosta, "Ambiguity resolution in direction of arrival estimation using mixed integer optimization and deep learning," in *Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON)*, pp. 317–322, Dayton, OH, USA, July 2019.
- [53] Q. Fu, B. Jing, and P. He, "Blind DOA estimation in a reverberant environment based on hybrid initialized multi-channel deep 2-D convolutional NMF with feedback mechanism," *IEEE Access*, vol. 7, pp. 179679–179689, 2019.
- [54] S. Chakrabarty and E. A. Habets, "Multi-speaker DOA estimation using deep convolutional networks trained with noise signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 8–21, 2019.
- [55] M. Wajid, B. Kumar, A. Goel, A. Kumar, and R. Bahl, "Direction of arrival estimation with uniform linear array based on recurrent neural network," in *Proceedings of the 2019 5th International Conference on Signal Processing, Computing And Control (ISPCC)*, pp. 361–365, Solan, India, October 2019.
- [56] Q. Pan, C. Mei, N. Tian, B. W.-K. Ling, and E. X. Wang, "Source enumeration based on a uniform circular array in a determined case," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 700–712, 2018.
- [57] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8549–8560, 2018.
- [58] Z.-M. Liu, C. Zhang, and P. S. Yu, "Direction-of-arrival estimation based on deep neural networks with robustness to array imperfections," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 12, pp. 7315–7327, 2018.
- [59] Y. Kase, T. Nishimura, T. Ohgane, Y. Ogawa, D. Kitayama, and Y. Kishiyama, "DOA Estimation of Two Targets with Deep Learning," in *Proceedings of the 2018 15th Workshop On Positioning, Navigation And Communications (WPNC)*, pp. 1–5, Bremen, Germany, October 2018.
- [60] D. Wang, Y. Zou, and W. Wang, "Learning soft mask with DNN and DNN-SVM for multi-speaker DOA estimation using an acoustic vector sensor," *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 1692–1709, 2018.
- [61] S. Chakrabarty and E. A. Habets, "Broadband DOA estimation using convolutional neural networks trained with noise signals," in *Proceedings of the 2017 IEEE Work-Shop on Applications of Signal Processing to Audio And Acoustics (WASPAA)*, pp. 136–140, New Paltz, NY, USA, October 2017.
- [62] W. Zheng, Y. X. Zou, and C. Ritz, "Spectral mask estimation using deep neural networks for inter-sensor data ratio model based robust DOA estimation," in *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 325–329, South Brisbane, Australia, April 2015.
- [63] Q. Li, X. Zhang, and H. Li, "Online direction of arrival estimation based on deep learning," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech And Signal Processing (ICASSP)*, pp. 2616–2620, Calgary, Canada, April 2018.