Received: 12 December 2017 | Revised: 15 February 2018 | Accepted: 26 February 2018

DOI: 10.1002/mmce.21282

RESEARCH ARTICLE

WILEY INTERNATIONAL JOURNAL OF RF AND MICROWAVE COMPUTER-AIDED ENGINEERING

Direction of arrival estimation for smart antenna in multipath environment using convolutional neural network

Youssef Harkouss | **Hassan Shraim** | **Hussein Bazzi**

Department of Electrical and Electronic Engineering, Faculty of Engineering III, Lebanese University, Hadath, Beirut, Lebanon

Correspondence
Youssef Harkouss, Faculty of Engineering III, Lebanese University, Hadath, Beirut, Lebanon.
Email: harkouss@ul.edu.lb

Abstract

This article aims to present a novel direction of arrival (DOA) estimation strategy for smart antenna in multipath environment. The smart antenna is composed of 2 main parts: the DOA estimator and the switched-beam system. In the first part, a DOA estimation method based on convolutional neural network (CNN) has been implemented. The CNN is capable to select the desired radiation beams of the switched-beam antenna without knowing the number of source signals coming from different directions, and in the case of noncoherent and coherent signals. Simulation results have been presented to show the effectiveness of the proposed intelligent approach.

KEYWORDS

antenna array, convolutional neural network, DOA estimation, switched-beam system

1 | INTRODUCTION

Wireless technology is one of the main areas of research in today's world of communication, and the study of wireless communication system is incomplete without the deep study of the operation of antennas. In recent years, smart antenna becomes a very hot research and can significantly improve the efficiency of wireless communication systems.[1,2] Direction of arrival (DOA) estimation strategy is the heart of the smart antenna. In general it plays an important role in array signal processing.[3] Estimation of angle of arrival of some signals is important for systems using adaptive antenna arrays and in interference environments. Smart antenna systems are capable of automatically forming beams in the directions of the desired signals and steering nulls in the direction of the interfering signals. Many algorithms have been proposed for the DOA estimation. Eigen-decomposition-based algorithms, including multiple signal classification (MUSIC)[4,5] and estimation of signal parameters via rotational invariance techniques (ESPRIT),[6,7] have high-resolution DOA estimation performance, but they fail to work in the case of coherent signals. Some smoothing techniques including spatial smoothing techniques,[8] subspace smoothing techniques,[9] and temporal smoothing techniques[10] were proposed to resolve this coherent problem. However these algorithms require extensive computation and are difficult to implement in real-time, meaning these methods are not suitable for real-time applications.

Neural networks have widely shown important performance in DOA estimation procedure.[11–14] They do not perform eigen-decomposition processes. They are found to be faster than the conventional super-resolution techniques. The major disadvantage of the neural network DOA estimation proposed in these references[11–14] is that the number of output neurons used in the output layer of the neural network is equal to the number of arriving signals on the antenna array.

In this article, a novel DOA estimation strategy-based on convolutional neural network (CNN) is investigated. This intelligent DOA estimation is implemented for a switched-beam antenna. The CNN is capable to select the desired radiation beams without knowing the number of source signals coming from different directions. The signal sources can be coherent or noncoherent. The resolution of MUSIC and ESPRIT is diminishes when signal sources are coherent.

Recently, CNN models have resulted in state of the art performance on a broad range of computer vision tasks such as face recognition,[15] large-scale object classification,[16] and document analysis and recognition.[17,18] In medical image analysis, CNNs have shown great potential for various

applications such as medical image pattern recognition[19] and tissue classification.[20]

A switched-beam antenna system produces a set of predefined beams. The most popular switched-beam system is the system based on Butler matrix.[14,21] In comparison with adaptive antenna array systems, the technique of this switched-beam system is simple and very appropriate for real-time processing applications.

This paper is structured as follows. Section 2 develops data model. Section 3 presents the structure of the CNN and its training algorithm. The switched-beam antenna system is described in Section 4. Section 5 describes the proposed smart antenna system. Section 6 presents different obtained results, and Section 7 summarizes our conclusions.

2 | DATA MODEL

A uniform linear array (ULA) with spacing d is considered in this paper. The structure of ULA is shown in Figure 1. This antenna array contains M elements. The received signal obtained at the output of this antenna array is given by:

$$x(t)=\mathrm{A}S(t)+n(t) \tag{1}$$

where $x(t)=(x_1(t), x_2(t), \ldots, x_M(t))^{\mathrm{T}}$. $S_1(t), S_2(t), \ldots, S_D(t)$ are the signals coming from the directions $\theta_1, \theta_2, \ldots, \theta_D$ respectively and $S(t)=(S_1(t), S_2(t), \ldots, S_D(t))^{\mathrm{T}}$. $n(t)=(n_1(t), n_2(t), \ldots, n_M(t))^{\mathrm{T}}$ is the received noise vector and A is the steering matrix. A can be written as:

$$\mathrm{A}=(a(\theta_1), a(\theta_2), \ldots, a(\theta_D))^{\mathrm{T}}$$

where $a(\theta_k)=\left(1, e^{-j\varphi_k}, e^{-j2\varphi_k}, \ldots, e^{-j(\mathrm{M}-1)\varphi_k}\right)^T$ is the steering vector and

$$\varphi_k=2\pi\frac{d}{\lambda}\sin(\theta_k)\ (k=1,\ldots,D) \tag{2}$$

A is then given by:

$$A=\begin{pmatrix} 1 & 1 & \ldots & 1 \\ e^{-j\varphi_1} & e^{-j\varphi_2} & \ldots & e^{-j\varphi_D} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-j(\mathrm{M}-1)\varphi_1} & e^{-j(\mathrm{M}-1)\varphi_2} & \ldots & e^{-j(\mathrm{M}-1)\varphi_D} \end{pmatrix} \tag{3}$$

The superscript T denotes the transpose and λ is the wavelength.

The output signal of the mth element is then:

$$x_m(t)=\sum_{k=1}^{D} s_k(t)e^{-j(m-1)\varphi_k}+n_m(t)\ (m=1,\ldots,M) \tag{4}$$

For array output $x(t)$, the covariance matrix R is given by:

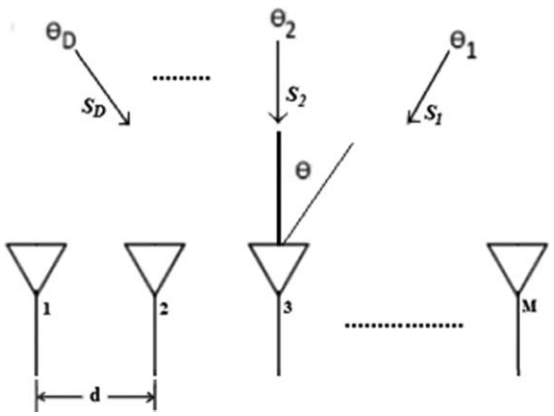

FIGURE 1 Uniform linear array

$$R=\frac{1}{K}\sum_{i=1}^{K} x(i)x^H(i) \tag{5}$$

where H denotes the conjugate transpose. K is the number of snapshots or the number of samples in the time domain and $x(i)=(x_1(i), x_2(i), \ldots, x_M(i))^{\mathrm{T}}$.

3 | CONVOLUTIONAL NEURAL NETWORK

Figure 2 shows the scheme of a CNN. This network is composed of $2r+2$ layers (r is a positive integer): $r+1$ convolution layers, r subsampling layers and a fully-connected layer. The different layers are arranged as follows: $C1$, $S2$, $C3$, $S4$, …, $S2r$, $C(2r+1)$, $F(2r+2)$ where Ci ($i=1, 3, 5, \ldots, 2r+1$) represents a convolution layer, Sj ($j=2, 4, 6, \ldots, 2r$) represents a subsampling layer and $F(2r+2)$ is the fully-connected output layer. Each convolution layer Ci ($i=1, 3, 5, \ldots, 2r-1$) and each subsampling layer Sj contains several planes, and each plane is a 2D array of neurons.

In a convolution layer Ci ($i=3, 5, \ldots, 2r+1$), each convolution plane CP_k^i ($k=1, 2, \ldots, n_i$) is connected to one or more subsampling planes (SPs) of the preceding subsampling layer $S(i-1)$. n_i is the number of planes in Ci. Note that each plane CP_k^1 ($k=1, 2, \ldots, n_1$) of $C1$ is connected directly to the input of the CNN. We associate to each connection of CP_k^i a matrix of weights CM_q^i ($q=1,2,\ldots,t_k^i$). This matrix is called convolution mask. t_k^i is the number of connections of CP_k^i. The output of CP_k^i is a matrix and is given by:

$$O_{\mathrm{CP}_k^i}=f_i^{cl}\left(\sum_{q=1}^{t_k^i} O_{\left(\mathrm{SP}^{i-1}\right)_q}\otimes CM_q^i+b_{\mathrm{CP}_k^i}\right) \tag{6}$$

$O_{\left(\mathrm{SP}^{i-1}\right)_q}$ is the output of the plane $\left(\mathrm{SP}^{i-1}\right)_q$ and $\left(\mathrm{SP}^{i-1}\right)_q$ is the SP of $S(i-1)$ which is connected to CP_k^i through the qth connection of CP_k^i. $b_{\mathrm{CP}_k^i}$ ($\in\mathbb{R}$) is the bias term of CP_k^i, $\otimes$ is the 2D convolution operator and f_i^{cl} is the activation function of Ci.

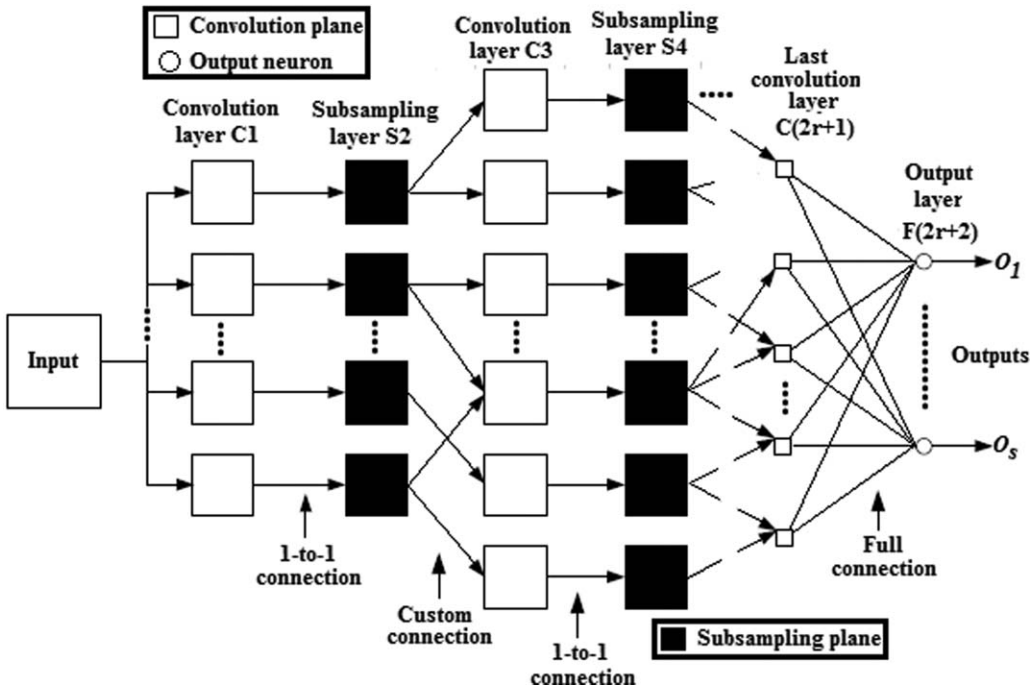

FIGURE 2 Structure of a CNN

In a subsampling layer Sj (j=2, 4, 6, …, $2r$), each SP S P_l^j (l=1, 2, …, m_j) is connected only to the plane CP_l^{j-1} of the preceding convolution layer $C(j-1)$. m_j is the number of planes in Sj. Note that Sj and $C(j-1)$ have the same number of planes ($m_j=n_{j-1}$). For the calculation of the output $O_{SP_l^j}$ of the plane SP_l^j, we isolated firstly each block 2 × 2 (or block of 4 pixels) of $O_{CP_l^{j-1}}$ (Figure 3) and then we used this block 2 × 2 to calculate the corresponding pixel of $O_{SP_l^j}$. The output of SP_l^j is then a matrix and is given by:

$$O_{SP_l^j}=f_j^{sl}\left(S\times w_{SP_l^j}+b_{SP_l^j}\right) \tag{7}$$

f_j^{sl} is the activation function of Sj, $b_{SP_l^j}$ ($\in \mathbb{R}$) is the bias term of SP_l^j and $w_{SP_l^j}$ ($\in \mathbb{R}$) is the weight associated to the connection $CP_l^{j-1} \rightarrow SP_l^j$. S is a matrix where each element of this matrix is the summation of 4 pixels in the block 2 × 2. Figure 3 shows the calculation of $O_{SP_l^j}$ using Equation 7.

In a convolution or subsampling layer all planes have the same size and in a convolution layer all convolution masks have also the same size. Table 1 shows the size of each plane (convolution and SPs) in CNN. In the last convolution layer $C(2r+1)$, the size of the convolution mask is $W^{2r}\times W^{2r}$. We can conclude that, the output of each plane in this layer is a scalar. In the output layer $F(2r+2)$, each neuron is connected to all planes of $C(2r+1)$. The output of the *ith* neuron in this layer (i=1, 2, …, s) is a scalar and is given by:

$$O_i=f^{ol}\left(\sum_{h=1}^{n_{2r+1}} O_{CP_h^{2r+1}}\times w_{ih}+b_i\right) \tag{8}$$

b_i ($\in \mathbb{R}$) is the bias term of this neuron, f^{ol} is the sigmoid activation function and w_{ih} is the weight of the connection C $P_h^{2r+1} \rightarrow ith$ output neuron in the output layer. $O_{CP_h^{2r+1}}$ is the output of the plane CP_h^{2r+1} and s is the number of neurons in the output layer $F(2r+2)$.

The network CNN is subject to a training procedure before its use in any environment. The aim of this training procedure is to minimize an error function, which is defined in terms of the network actual outputs and the network desired outputs. In this article, the training algorithm used to train the CNN is the resilient back-propagation RPROP.[22] This algorithm is one of the fastest among the first order training algorithms (BackPropagation, BackPropagation with Momentum and Variable learning rate, and Super Self-Adapting Backpropagation) and it ensures easily the requested high accuracy for the classification. First order training algorithm is the recommended algorithm when we have big neural networks and large training data sets.

4 | SWITCHED-BEAM ANTENNA SYSTEM

A switched-beam antenna system produces a set of predefined beams. In this system, a switch is used to select the

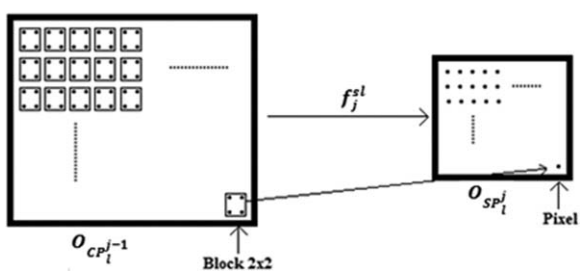

FIGURE 3 Calculation of $O_{SP_l^j}$

TABLE 1 Size of each 2D array in CNN

Input of CNN	Plane CP_k^i (i=1, 3, 5, …, 2r+1) (k=1, 2, …, n_i) Size of convolution mask *in* $Ci = W_{CM^i}$	Plane SP_l^j (j=2, 4, 6, …, 2r) (l=1, 2, …, m_j)
$W^0 \times W^0$	$W^i \times W^i$ $W^i = W^{i-1} - W_{CM^i} + 1$	$W^j \times W^j$ $W^j = W^{j-1}/2$

best beam(s) of receiving signals from a number of predefined beams (desired beams). Such systems offer many of the advantages of the fully adaptive systems at less expense and complexity. Figure 4 shows the switched-beam system used in this article. This system is based on Butler matrix.

4.1 | Microstrip patch antenna array

The proposed switched-beam system contains a linear array with M=16 antenna elements where d=λ/2. The smart antenna system operates at f=2.4 GHz. In this paper, the elements are rectangular microstrip patches because of their low cost, robustness and ease of implementation. The Rogers RT/duroid 5880 material of dielectric constant ε_r=2.2 has been used as the substrate of the antenna elements (h=3.2 mm, L=40.3074 mm, and W=49.4106 mm).

Each antenna element is excited using an inset microstrip line in order to fix an input impedance of 50 Ω. This type of feeding produces the linear polarization of the antenna element. Our work does not take into account the mutual coupling between the microstrip antennas. This is based on the fact that due to the mutual coupling, the radiation pattern of an antenna array will not lose its main emission/reception shape.

4.2 | Butler matrix

The Butler matrix is a microwave network, normally employed in beamforming and scanning networks for linear and circular antenna arrays. A Butler matrix in its standard form is a $M \times M$ network, which consists of hybrid couplers, phase shifters and crossovers. M is the number of input/output ports that give a set of M different beams. The Butler matrix is a reciprocal passive circuit. It is a parallel system which is composed of several junctions connecting input to output ports by transmission lines.

A $M \times M$ Butler matrix contains $(M/2)\log_2(M)$ couplers, $(M/2)(\log_2(M)-1)$ phase shifters and $\sum_{k=1}^{\log_2(M)} \left[\frac{M}{2}\left(2^{k-1}-1\right)\right]$ crossovers. The Butler matrix is composed of several stages of hybrid couplers (3 dB, 90°). Each coupler, if it is balanced, divides the input signal into 2 signals. If we consider that we have n (n is an integer) stages of couplers, and we need to obtain a signal with uniform distribution in amplitude, we will have 2^n outputs per input. Number of output ports will be equal to that of input ports (2^n), and this due to the symmetry of directional couplers. Butler matrix

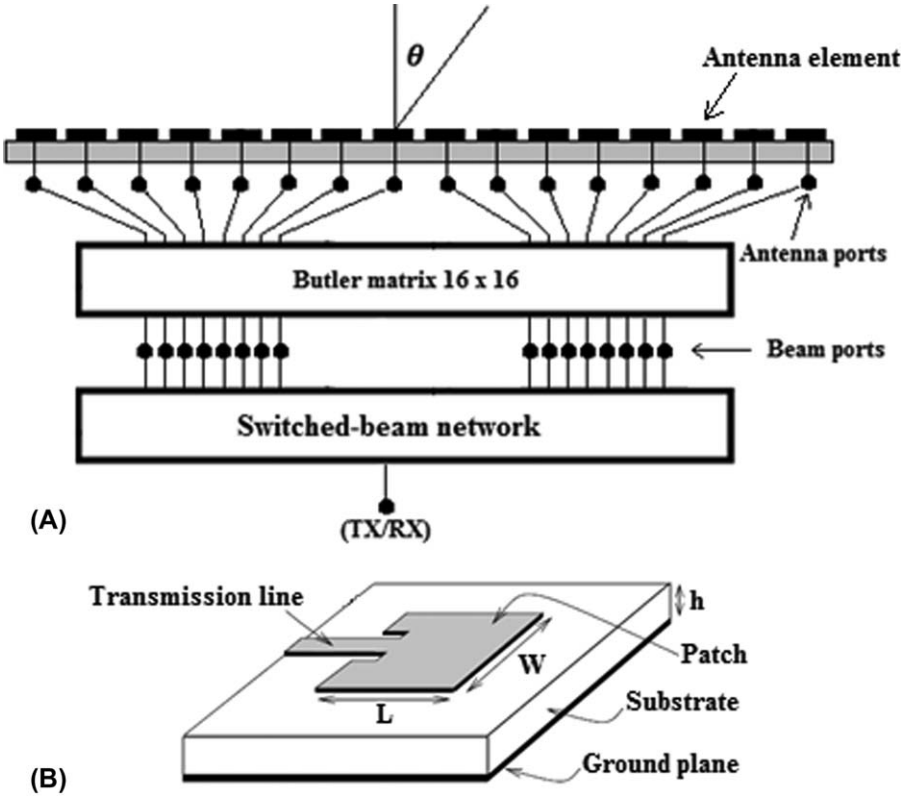

FIGURE 4 (A) Structure of the switched-beam antenna system with 16 antenna elements, (B) Antenna element

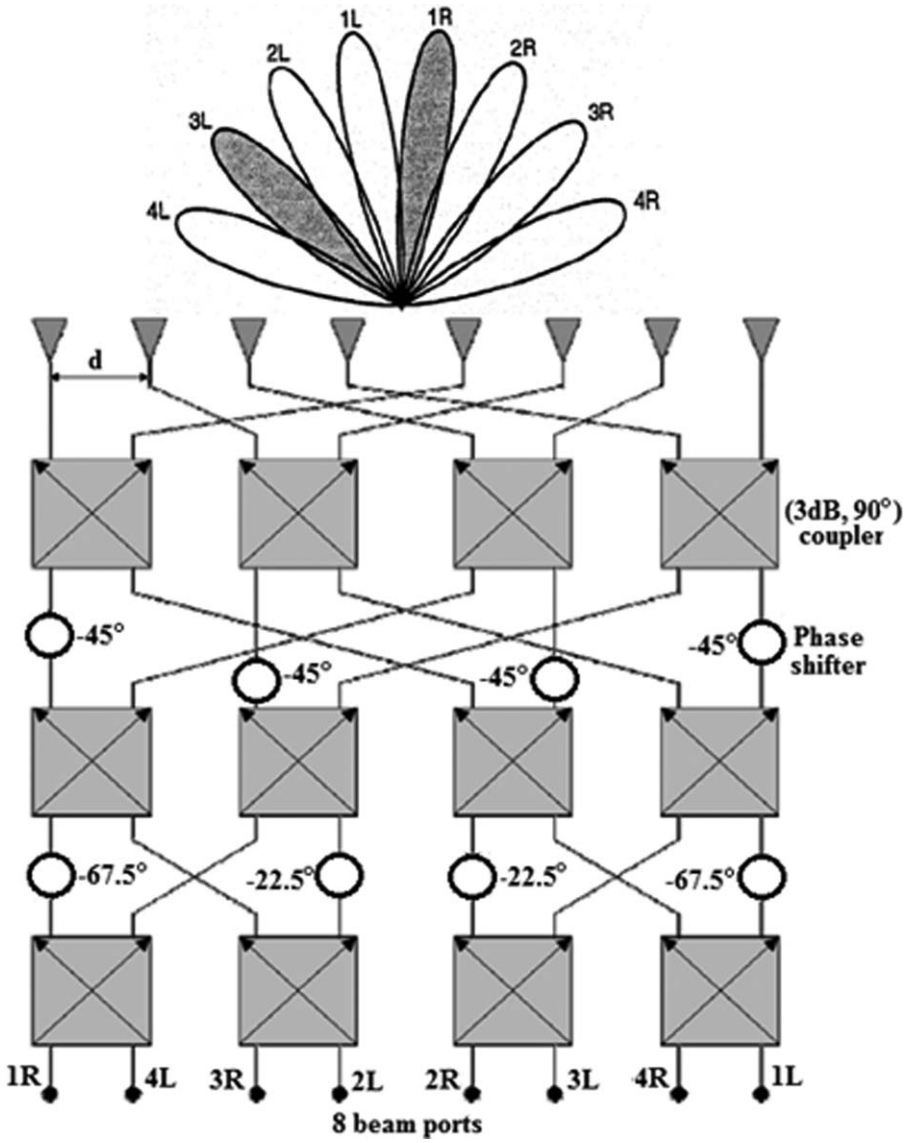

FIGURE 5 8 × 8 Butler matrix block diagram

with hybrid couplers (3 dB, 90°) is a symmetrical circuit. In the theory of transmission lines, each displacement of a wave in a transmission line generates a phase which is related to the covered distance. That means, each transmission line may be considered as an electromagnetic wave phase shifter, and the obtained phase depends on the length of the transmission line at the operation frequency. Phase shifters are designed based on this principle by using transmission lines with microstrip, coplanar, or other technologies. On the other hand, a crossover is placed in the intersection of transmission lines to prevent signals combinations.

A standard Butler matrix is then denoted as a $M \times M$ system where $M=2^n$. The number of hybrid couplers is $Mn/2$, that are arranged in n ranges. In each range, $M/2$ couplers are found. The number of ranges of phase shifters is $n-1$. The number and the position of phase shifters depend on the type of hybrid couplers used. If 90° couplers are used, then the number of phase shifters per range is $M/2$. Figure 5 shows the block diagram of a $M \times M$ Butler matrix where $M=8$ and the desired beams generated by this matrix: 4L, 3L, 2L, 1L, 1R, 2R, 3R, and 4R.

When a signal is injected in a beam port of the Butler matrix, a fixed phase difference is produced between different output ports. Each phase difference produces a beam deviation with respect to the normal of the array axis. Each input port can be then creating a beam in a different direction in space. M different beams can thus be obtained. The different angles of these M beams are given by:

$$\theta_k = \arcsin\left(\frac{\lambda \psi_k}{2\pi d}\right) \quad (k \in [1, M/2])$$

ψ_k is the phase gradient between 2 consecutive radiated elements where $\psi_k = \pm(2k-1)\frac{\pi}{M}$, θ_k is the angle between the

TABLE 2 Beam directions for a 16-elements antenna array

Phase difference (deg.)	Angle θ (deg.) of the beam maximum radiation (angle with the normal of the antenna array)	Beam
–168.75	–69.64	8L
–146.25	–54.34	7L
–123.75	–43.43	6L
–101.25	–34.23	5L
–78.75	–25.94	4L
–56.25	–18.21	3L
–33.75	–10.81	2L
–11.25	–3.58	1L
11.25	3.58	1R
33.75	10.81	2R
56.25	18.21	3R
78.75	25.94	4R
101.25	34.23	5R
123.75	43.43	6R
146.25	54.34	7R
168.75	69.64	8R

beam k (maximum radiation) and the normal to the axis of the array, d is the distance between 2 consecutive radiated elements and λ is the wavelength. Note that each value of k gives 2 values of θ_k, that is, angles of beams kL and kR.

Butler Matrix ensures then 2 functions: (1) distribution of RF signals to different elements of the array, and (2) direction and orthogonality of beams. Butler matrix generates then orthogonal beams (optimal combination) and its design becomes easier thanks to the symmetry of its topology.

By connecting Butler matrix to an antenna array and RF switched-beam network, multiple beams can be then formed. By introducing an RF signal to a beam port of the 8 × 8 Butler matrix, and via progressive phases, we can form a beam in a certain direction of space. In a similar manner, if we introduce a second RF signal, another beam will be formed in another direction. If the 2 beam ports 1R and 3L of the 8 × 8 Butler matrix are also excited at the same time, the 2 beams 1R and 3L will be produced simultaneously in 2 different directions (Figure 5).[23] In certain cases, there is a need to create additional beams which are normally generated by Butler matrix. For example, in a 8 × 8 Butler matrix, several other beams in addition to existing beams can be created by combining 2 beam ports of this matrix: (1) beam port 1L and beam port 1R to create a beam on the normal of the array, and (2) beam port 2R and beam port 3R to create a beam between the 2 beams 2R and 3R.

The feeding network used in this paper is a 16 × 16 Butler matrix which is a fixed network of 16 input ports (beam ports) and 16 output ports (antenna ports) (Figure 4). The Butler matrix contains 32 hybrid couplers (3 dB, 90°) and 12 phase shifters. The beam ports are connected to a switching network that is responsible for the excitation of the desired beam(s). Table 2 shows beam directions of antenna elements fed by this 16 × 16 Butler matrix. These 16 orthogonal predefined beams are produced by uniform illumination of 16 microstrip patches (Figure 7A).

5 | PROPOSED SMART ANTENNA SYSTEM

Figure 6 shows the structure of the proposed smart antenna system using CNN.

The proposed system has 2 parts: the DOA estimator using CNN and the switched-beam antenna system. Block A gives firstly the output signal $x(t)$ (Equation 1). Block B calculates then the covariance matrix R. Block C constructs the matrix I using the real part and the imaginary part of the

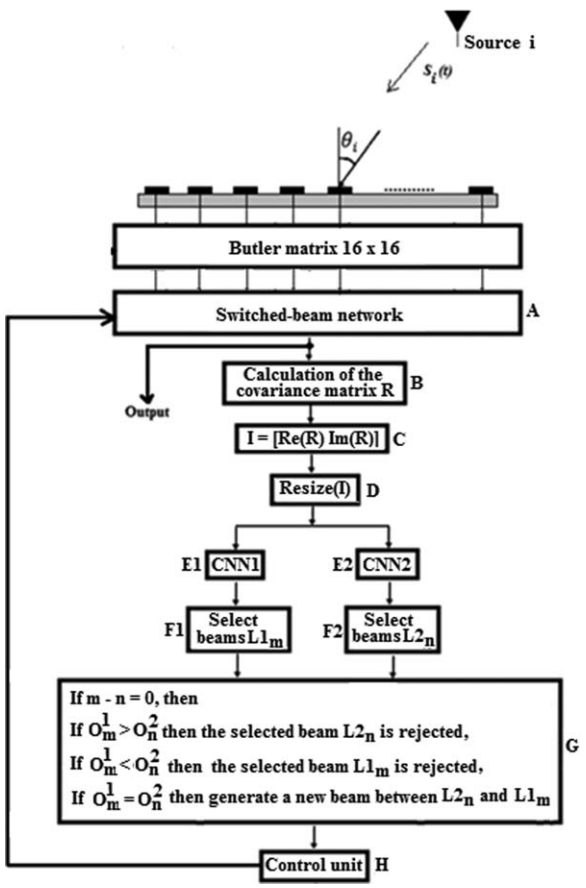

FIGURE 6 Block diagram of the proposed smart antenna system using CNN

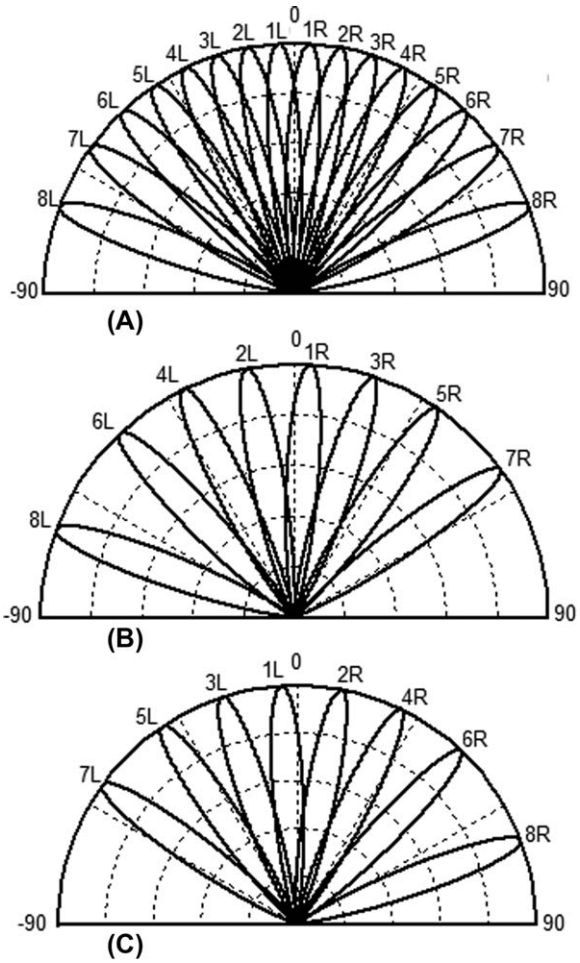

FIGURE 7 Desired beams: (A) 16 orthogonal beams produced by the uniform illumination of 16 microstrip patches fed by an 16 × 16 Butler Matrix, (B) Desired beams for CNN1 (8 beams), (C) Desired beams for CNN2 (8 beams)

matrix R. I is a matrix of dimension $M\times 2M$ where M is the number of array elements. Block D transforms the matrix I to an image I and makes a resizing of this image whose size becomes $2M\times 2M$. The resizing is done in Matlab by using bicubic interpolation method. This resizing process is obligatory for the CNN network since this network receives at its input only square images. The obtained image I will be applied to the block E1 (CNN1) and the block E2 (CNN2). The network CNNl ($l=1,2$) gives at the output a vector $O^l=\left(O_1^l,O_2^l,\ldots,O_8^l\right)^T$. Each CNN$l$ has 8 neurons in the output layer (see Section 6.2), 1 neuron for each beam. Note that, if the incident signal arrives to a beam of a network CNNl, then the corresponding output of this network must be greater than or equal to 0.5. This vector will be transformed into another vector $O^{bl}=\left(O_1^{bl},O_2^{bl},\ldots,O_8^{bl}\right)^{\mathrm{T}}$ (b: binary) which contains only 0s and/or 1s (see Section 6.4).

The 2 blocks F1 and F2 select the beams corresponding to the input I applied to E1 and E2. F1 selects the beams $L1_m$ ($m\in\{1,2,3,4,5,6,7,8\}$) where $\{L1_1=8L, L1_2=6L, L1_3=4L,\ L1_4=2L, L1_5=1R, L1_6=3R, L1_7=5R, L1_8=7R\}$ are the desired beams of CNN1 (Figure 7B). The HPBW (half power beam width) of each beam of CNN1 is approximately equal to 15°. F2 selects the beams $L2_n$ ($n\in\{1,2,3,4,5,6,7,8\}$) where $\{L2_1=7L, L2_2=5L, L2_3=3L, L2_4=1L, L2_5=2R, L2_6=4R, L2_7=6R, L2_8=8R\}$ are also the desired beams of CNN2 (Figure 7C). The HPBW of each beam of CNN2 is also approximately equal to 15°. Figure 7 shows the direction of the maximum radiation of each beam. For example, if $O^{b1}=(1\ 0\ 0\ 1\ 0\ 0\ 0\ 0)^{\mathrm{T}}$ and $O^{b2}=(0\ 0\ 0\ 0\ 1\ 1\ 0\ 0)^{\mathrm{T}}$, then $L1_m\in\{L1_1,L1_4\}$ and $m\in\{1,4\}$ (beams $8L, 2L$ are selected by CNN1), and $L2_n\in\{L2_5,L2_6\}$ and $n\in\{5,6\}$ (beams $2R, 4R$ are selected by CNN2). If there are 2 consecutive beams at the input of block G or if m-n = 0 at the input of this block, then: (1) if $O_m^1=O_n^2$, i.e. $0.49\le O_m^1$ and $O_n^2\le 0.55$, then generate a new beam between $L2_n$ and $L1_m$ by combining the 2 beam ports corresponding to these 2 beams, (2) if $O_m^1>O_n^2$ (O_m^1 and $O_n^2>0.55$) then the corresponding selected beam $L2_n$ is rejected, and (3) if $O_m^1<O_n^2$ (O_m^1 and $O_n^2>0.55$) then the corresponding selected beam $L1_m$ is rejected. Note that O_m^1 is the mth output of the network CNN1 and O_n^2 is the nth output of the network CNN2. The control unit (block H) controls the switched-beam network and then the generation of the desired radiation pattern of the smart antenna. This radiation pattern is formed using the selected beams $L1_m$ and $L2_n$.

6 | CNN FOR DOA ESTIMATION AND RESULTS

6.1 | Training and test data sets

The covariance matrices are used to train and test each CNN. The networks are then trained by examples generated from Equation 5. In this case, $N=18524$ training examples that we refer to as the training data set, are used to train each CNN network. Note that the CNN architectures are capable of learning large data set. The training data set of each CNNl ($l=1,2$) is given by:

$$S_{\text{train}}^l=\left\{\left(I^{li},O_d^{bli}\right)/i=1,\ldots,N\right\} \tag{9}$$

where $\left(I^{li},O_d^{bli}\right)$ is the ith training example. I^{li} is the ith input of CNNl and O_d^{bli} is the ith corresponding desired output

TABLE 3 Parameter values used to generate training and test data sets

Number of array elements	$M=16$
Number of snapshots (K)	[10, 200]
Distance between 2 consecutive array elements	λ/2 (λ: carrier wavelength)
SNR	[0, 10 dB]

TABLE 4 Narrow band signals used to train and test CNN1 network

Number of narrow band signals	Incident angles
1	$\theta \in \boldsymbol{S}_1$
2 (noncoherent signals)	$\theta_1, \theta_2 \in \boldsymbol{S}_1$
2 (coherent signals)	$\theta_1, \theta_2 \in \boldsymbol{S}_1$
2 (noncoherent signals)	$\theta_1 \in \boldsymbol{S}_1$ and $\theta_2 \in \boldsymbol{S}_2$
2 (coherent signals)	$\theta_1 \in \boldsymbol{S}_1$ and $\theta_2 \in \boldsymbol{S}_2$

vector. For example, if I is the matrix generated by 2 arriving signals $S_1$and S_2 where the angles of arrival are θ_1 (θ_1 belongs 100% to $8L$) and θ_2 (θ_2 belongs 100% to $3L$), respectively, then the corresponding desired output of CNN1 and CNN2 are $O_d^{b1} = (1\ 0\ 0\ 0\ 0\ 0\ 0\ 0)^T$ and $O_d^{b2} = (0\ 0\ 1\ 0\ 0\ 0\ 0\ 0)^T$ respectively.

Tables 3–5 list different parameters and incident angles adopted to generate the training and test data sets. The sets $\boldsymbol{S}_1 = \{8L, 6L, 4L, 2L, 1R, 3R, 5R, 7R\}$ and $\boldsymbol{S}_2 = \{7L, 5L, 3L, 1L, 2R, 4R, 6R, 8R\}$ contain the desired beams of CNN1 and CNN2 respectively. 4 values of SNR (signal-to-noise ratio) (0, 3, 7 and 10 dB) and 4 values of K (10, 50, 100 and 200) are used to generate the training and test data sets. Each CNN network is trained and tested by both noncoherent (Figure 1) and coherent (Figure 8) signals. Figure 8 shows an example of direct and multipath signals coming from the same source i and impinging on the antenna array. The problem involving coherent signals is a fatal problem for subspace algorithms. It will greatly affect the performance of DOA estimation. The association of the CNN-based DOA estimation method and the switched-beam antenna can be then a good solution for smart antenna systems in multipath environment and it is very suitable for real-time processing applications.

The obtained networks are then tested on sets of data that we refer to as the test data sets. Each test data set contains 2304 examples. These sets are formed by examples not used to train CNN networks where SNR = 4 dB and K = 150. We consider the cases of 2 noncoherent signals. Testing identified CNN networks on a set of 'fresh' data is called validation process. It is a commonly used method for evaluating the performance of identified CNN networks.

TABLE 5 Narrow band signals used to train and test CNN2 network

Number of narrow band signals	Incident angles
1	$\theta \in \boldsymbol{S}_2$
2 (noncoherent signals)	$\theta_1, \theta_2 \in \boldsymbol{S}_2$
2 (coherent signals)	$\theta_1, \theta_2 \in \boldsymbol{S}_2$
2 (noncoherent signals)	$\theta_1 \in \boldsymbol{S}_1$ and $\theta_2 \in \boldsymbol{S}_2$
2 (coherent signals)	$\theta_1 \in \boldsymbol{S}_1$ and $\theta_2 \in \boldsymbol{S}_2$

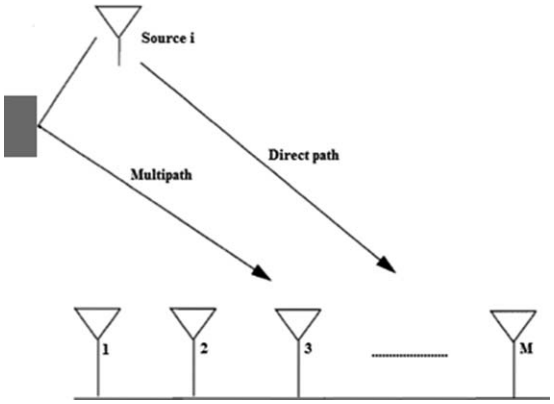

FIGURE 8 Direct and multipath signals from source i

6.2 | CNN1 and CNN2 structure

In this section, we are seeking to find the CNN network that gives the highest performance. Thus, we will study the performance of several different architectures of the CNN network by varying the number of features in each layer, the connection matrix from the subsampling to the convolution layer and the activation function used in each convolution or subsampling layer. The obtained results show that among the different architectures trained and tested, the architecture presented in this section scored the highest accuracy in training and testing processes respectively. Each proposed CNN network is then composed of 3 convolution layers (C1, C3, and C5), 2 subsampling layers (S2 and S4) and a fully connected output layer (F6). Both networks CNN1 and CNN2 have then the same structure. Table 6 shows the topology of the proposed networks CNN1 and CNN2, and the connections between different convolution and SPs. The input of each CNNl (l=1, 2) is the 32×32 matrix I and the output vector is $O^l = (O_1^l, O_2^l, \ldots, O_8^l)^{\mathrm{T}}$ for each network. CNNl has then 8 neurons in the output layer. CP_k^i is the kth CP in the ith

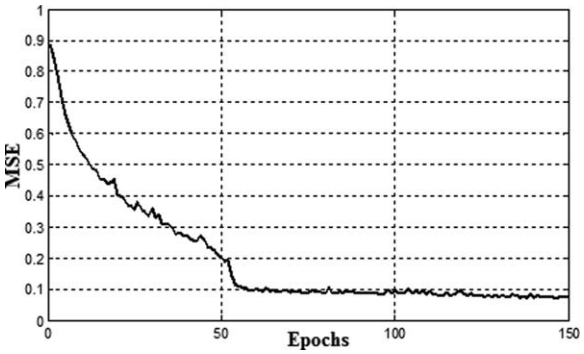

FIGURE 9 MSE as a function of the number of epochs for the network CNN1

TABLE 6 CNN1 and CNN2 structure ($SP_1^4 \rightarrow CP_1^5$, CP_2^5, CP_3^5 means that SP_1^4 is connected to CP_1^5, CP_2^5, and CP_3^5)

Layer	C1	S2	C3	S4	C5
Number of CPs and SPs)	2 CP_1^1, CP_2^1	2 SP_1^2, SP_2^2	5 CP_1^3, CP_2^3, CP_3^3, CP_4^3, CP_5^3	5 SP_1^4, SP_2^4, SP_3^4, SP_4^4, SP_5^4	6 CP_1^5, CP_2^5, CP_3^5, CP_4^5, CP_5^5, CP_6^5
Convolution mask	5×5		3×3		6×6
Connections	$CP_1^1 \rightarrow SP_1^2$ $CP_2^1 \rightarrow SP_2^2$	$SP_1^2 \rightarrow CP_1^3$, CP_4^3, CP_5^3 $SP_2^2 \rightarrow CP_2^3$, CP_3^3	$CP_1^3 \rightarrow SP_1^4$ $CP_2^3 \rightarrow SP_2^4$ $CP_3^3 \rightarrow SP_3^4$ $CP_4^3 \rightarrow SP_4^4$ $CP_5^3 \rightarrow SP_5^4$	$SP_1^4 \rightarrow CP_1^5$, CP_2^5, CP_3^5 $SP_2^4 \rightarrow CP_2^5$, CP_3^5, CP_4^5 $SP_3^4 \rightarrow CP_3^5$, CP_4^5, CP_5^5 $SP_4^4 \rightarrow CP_4^5$, CP_5^5, CP_6^5 $SP_5^4 \rightarrow CP_1^5$, CP_5^5, CP_6^5	CP_i^5 is connected to all output neurons ($i = 1, \ldots, 6$)

convolution layer and SP_k^j is the kth SP in the jth subsampling layer. The activation functions used in each convolution layer and subsampling layer are sigmoid function and linear function respectively.

6.3 | Training and test procedures of CNN networks

Each CNN is trained with the RPROP algorithm in the direction to minimize the mean squared error (MSE) between the network output vector O^{li} and the desired output vector O_d^{bli}:

$$\mathrm{MSE}_{\mathrm{CNN}l} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{s}\left(O_k^{li} - O_{dk}^{bli}\right)^2 \ (l = 1, 2) \qquad (10)$$

where N is the number of examples in the training data set. $O_d^{bli} = \left(O_{d1}^{bli}, O_{d2}^{bli}, \ldots, O_{ds}^{bli}\right)^{\mathrm{T}}$ is the ith desired output of the network CNNl and $O^{li} = \left(O_1^{li}, O_2^{li}, \ldots, O_s^{li}\right)^T$ is the ith CNNl output. Note that (I^{li}, O_d^{bli}) is the ith training example of CNNl.

In Figure 9, the MSE of the network CNN1 is plotted as a function of the number of epochs. During the training process and at each epoch, the MSE is calculated by Equation 10. The errors MSE calculated on the training data set and after the training process for the 2 networks CNN1 and CNN2 are: $\mathrm{MSE}_{\mathrm{CNN1train}} = 0.003101$ and $\mathrm{MSE}_{\mathrm{CNN2train}} = 0.00292$. A CNN network that is designed to generalize well will produce a correct input-output mapping, even when the input is different from the training examples used in the training data set. The errors MSE calculated on the test data set and after the training process are: $\mathrm{MSE}_{\mathrm{CNN1test}} = 0.00453$ and $\mathrm{MSE}_{\mathrm{CNN2test}} = 0.00412$.

6.4 | Results

In this section the results of simulated DOA estimation tests using the proposed intelligent approach are presented. The simulation procedure has been performed using MATLAB.

The performance of each CNN can be evaluated by visualizing 2 performances metrics: the error MSE and the accuracy which represents the degree of closeness of measurements of a quantity to that quantity's true value, defined as:

$$\mathrm{Accuracy}(\%) = \frac{N_{\mathrm{T}}}{N_{\mathrm{S}}} \times 100 \qquad (11)$$

where N_{S} is the total number of examples in the considered set (training data set, test data set, or any data set) and N_{T} is the

TABLE 7 Case of one arriving signal (SNR = 3 dB and $K = 50$)

Incident angle (θ)	Output vectors of CNN1 and CNN2 (real values)	Selected beam
19°	$O^1 = [0.0026\ -0.0166\ -0.0124\ -0.0001\ 0.0118\ \mathbf{0.9641}\ 0.0185\ 0.0001]^{\mathrm{T}}$ $O^2 = [0.0081\ -0.0055\ 0.0048\ 0.0061\ 0.0009\ -0.0414\ 0.011\ -0.013]^{\mathrm{T}}$	$m=6$ **Beam 3R**
–35°	$O^1 = [-0.0103\ 0.012\ 0.0237\ -0.026\ -0.001\ -0.0055\ 0.0011\ 0.0011]^{\mathrm{T}}$ $O^2 = [0.0099\ \mathbf{0.9746}\ -0.0034\ -0.0027\ 0.0069\ 0.0061\ -0.0044\ 0.0007]^{\mathrm{T}}$	$n=2$ **Beam 5L**
1°	$O^1 = [-0.0083\ 0.0106\ 0.0042\ 0.0179\ \mathbf{0.8877}\ -0.0029\ -0.0026\ 0.0026]^{\mathrm{T}}$ $O^2 = [-0.0016\ -0.0008\ 0.0006\ \mathbf{0.6542}\ 0.0049\ -0.0084\ 0.0073\ 0.0033\ ]^{\mathrm{T}}$	$m=5$ **Beam 1R** **(Beam 1L is rejected)**
–39.1°	$O^1 = [-0.0083\ \mathbf{0.5686}\ 0.0042\ 0.0179\ 0.0077\ -0.0029\ -0.0026\ 0.0026]^{\mathrm{T}}$ $O^2 = [-0.0016\ \mathbf{0.8208}\ 0.0006\ 0.0142\ 0.0049\ -0.0084\ 0.0073\ 0.0033\ ]^{\mathrm{T}}$	$n=2$ **Beam 5L** **(Beam 6L is rejected)**

TABLE 8 Case of 2 arriving signals with SNR = 7 dB and $K = 200$

Incident angles (θ_1, θ_2)	Output vectors of CNN1 and CNN2 (real values)	Selected beam(s)
–19°, 26° (NCS)	$O^1 = [0.0269\ 0.086\ 0.0265\ 0.0011\ 0.011\ 0.0283\ -0.0185\ 0.0142\ ]^T$ $O^2 = [0.009\ 0.0034\ \mathbf{0.9987}\ -0.0135\ 0.0052\ \mathbf{1}\ -0.0055\ 0.0001]^T$	$n=3, 6$ **Beams: 3L, 4R**
–11°, 70.2° (NCS)	$O^1 = [0.0194\ -0.0169\ 0.0296\ \mathbf{0.9861}\ 0.081\ 0.023\ 0.0022\ 0.0252]^T$ $O^2 = [0.013\ -0.0173\ 0.0109\ 0.0109\ 0.0185\ 0.01\ 0.0259\ \mathbf{0.9322}]^T$	$m=4, n=8$ **Beams: 2L, 8R**
35.2°, 35.2° (NCS)	$O^1 = [-0.0067\ 0.0143\ 0.04\ -0.0092\ 0.0035\ 0.009\ \mathbf{0.9259}\ 0.0153\ ]^T$ $O^2 = [-0.075\ 0.0231\ 0.016\ -0.026\ 0.011\ 0.0116\ 0.062\ 0.0131\ ]^T$	$m=7$ **Beam 5R**
–24.2°, 10.3° (CS)	$O^1 = [-0.0152\ 0.04\ \mathbf{0.9334}\ -0.0096\ -0.033\ 0.0293\ 0.027\ 0.058]^T$ $O^2 = [-0.0012\ 0.0021\ 0.0032\ 0.0112\ \mathbf{0.9976}\ 0.011\ -0.0401\ 0.02\ ]^T$	$m=3, n=5$ **Beams: 4L, 2R**
–11°, 38.4° (CS)	$O^1 = [-0.0152\ 0.04\ 0.0334\ \mathbf{0.9496}\ -0.033\ 0.0293\ \mathbf{0.827}\ 0.058]^T$ $O^2 = [-0.0012\ 0.0021\ 0.0032\ 0.0112\ 0.0976\ 0.011\ \mathbf{0.6103}\ 0.02\ ]^T$	$m=4, 7$ **Beams: 2L, 5R** **(Beam 6R is rejected)**

Abbreviation: NCS, noncoherent signals; CS, coherent signals.

The bold text represents the incident angles used to calculate different results, the significant obtained results in the output vectors of CNN1 and CNN2, and the best selected beams.

number of the correctly classified examples by the CNN. The accuracy of the CNN1 (and CNN2) calculated on the training data set is 100%. This accuracy reached 99.39% if the set of examples was the test data set. The performance of the CNN-based DOA estimation system is tested for both noncoherent and coherent signals. Tables 7–8 list different results obtained at the output of each CNN (real values) by varying the number of arriving signals on the antenna array and by considering the case of coherent and noncoherent signals. The incident angles of these results are different from the angles used to generate the training and test data sets. Table 7 shows the results obtained in the case of one arriving signal. Table 8 contains the results in the case of 2 coherent and noncoherent arriving signals.

For the selection of suitable beam(s), the output vector $O^l = \left(O_1^l, O_2^l, \ldots, O_8^l\right)^T$ ($l=1, 2$) of each CNNl (real values) will be transformed into another binary vector $O^{bl} = \left(O_1^{bl}, O_2^{bl}, \ldots, O_8^{bl}\right)^T$ which contains only 0s and/or 1s:

$$O_j^{bl} = \begin{cases} 0 \ \ if \ \ O_j^l < 0.5 \\ 1 \ \ otherwise \end{cases} \quad (j = 1, 2, \ldots, 8)$$

In the case of CNN1, if $O_j^{b1} = 1$ then the corresponding beam $L1_j$ is selected. For the network CNN2, if $O_j^{b2} = 1$, then the corresponding beam $L2_j$ is selected.

These results have demonstrated the efficiency and the ability of using a CNN for the DOA estimation in multipath environment, and proved that the CNN is able to generalize and to predict the output network for inputs never used in the training data set. The efficient CNN, once constructed, operates with a computation time which is negligible compared to the traditional techniques, such as MUSIC and ESPRIT algorithms. Due to the speed of CNNs, real-time applications can be then easily served into existing base stations.

7 | CONCLUSION

In this article, a novel and efficient DOA estimation strategy has been proposed for smart antenna system in multipath environment. This smart system is based on CNN and switched-beam system. The main advantage of this proposed intelligent system is that this system is capable to select suitable beams from the set of predefined beams even if the number of signal sources was unknown and variable. The number of each CNN outputs is independent of the number of arriving signals on antenna array. The proposed DOA estimation can be a good solution for smart antenna systems in multipath environment and is very suitable for real-time processing. Validations and experimental results will be an important perspective that we will work on to demonstrate the applicability and efficiency of CNN for the selection of suitable beams of the switched-beam antenna system

ORCID

Youssef Harkouss http://orcid.org/0000-0002-4765-7777

REFERENCES

[1] Stutzman WL, Reed JH, Dietrich CB, Kim BK, Sweeney DG. Recent results from smart antenna experiments-base station and handheld terminals. In IEEE Radio and Wireless Conference, RAWCON 2000; September, **2000**; Denver, USA, pp. 139–142.

[2] Kiong TS, Ismail M, Hassan A. WCDMA forward link capacity improvement by using adaptive antenna with genetic algorithm assisted MDPC beamforming technique. *J Appl Sci.* **2006**;6(8): 1766–1773.

[3] Hudson JE. *Adaptive Array Principles*. London: IEE Press;, **1989**.

[4] Lavate TB. Performance analysis of MUSIC and ESPRIT DOA estimation used in adaptive array smart antenna. *Int J Comput Netw.* **2010**;2(3):152–158.

[5] PaulrajOttersten A, Roy B, Swiindlehurst R, Xu AG, Kailath T. Subspace methods for direction-of arrival estimation. In: Bose, N.K., and Rao, C.R. eds. *Handbook of Statistics 10-Signal Processing and its Applications*; **1993**:693–640.

[6] Roy R, Kailath T. ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans Acoust Speech Sig Process.* **1989**;37(7):984–995.

[7] Han FM, Zhang XD. An ESPRIT-like algorithm for coherent DOA estimation. *IEEE Antennas Wirel Propag Lett.* **2005**;4(1): 443–446.

[8] Changuel H, Harabi F, Gharsallah A. 2-L-shape twodimensional arrival angle estimation with a classical subspace algorithm. *Prog Electromagn Res.* **2006**;66:301–315.

[9] Jeng SS, Lin HP, Okamoto G, Xu G, Vogel WJ. Multipathdirection finding with subspace smoothing. In *IEEE International Conference of Acoustics, Speech, and Signal Processing*, Munich, Germany; **1997**:3485–3488.

[10] Jiang H, Wang SX. On temporal smoothing for two dimensional direction-of-arrival estimation of coherent signals in multiplicative/additive noises environment. *Asia Pacific Conference on Communications*; **2005**:119–123.

[11] Du KL, Lai AKY, Cheng KKM, Swamy MNS. Neural methods for antenna array signal processing: A review. *Sig Process.* **2002**;82(4):547–561.

[12] Zooghby AH, Christodoulou CG, Georgiopoulos M. A neural network-based smart antenna for multiple source tracking. *IEEE Trans Antennas Propag.* **2000**;48(5):768–776.

[13] Mochida E, Iiguni Y. Adaptive DOA estimation using a radial basis function network. *Electron Commun Japan.* **2005**;88(9):11–20.

[14] Gotsis K, Siakavara AK, Sahalos JN. On the direction of arrival (DoA) estimation for a switched-beam antenna system using neural networks. *IEEE Trans Antennas Propag.* **2009**;57(5): 1399–1411.

[15] Lawrence S, Giles CL, Tsoi AC, Back AD. Face recognition: A convolutional neural network approach. *IEEE Trans Neural Netw.* **1997**;8(1):98–113.

[16] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inform Process Syst.* **2012**;25:1097–1105.

[17] Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), **2003**.

[18] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* **1998**;86(11): 2278–2324.

[19] Lo SB, Lou SA, Lin JS, Freedman MT, Chien MV, Mun SK. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Trans Med Imaging.* **1995**;14(4):711–718.

[20] Haj−Hassan H, Chaddad A, Harkouss Y, Desrosiers C, Toews M, Tanougast C. Classifications of multispectral colorectal cancer tissues using convolution neural network. *J Pathol Inform.* **2017**;8(1):1–7.

[21] Liberti J, Rappaport T. *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Application*. 1st ed. New Jersey: Prentice Hall PTR; **1999**.

[22] Riedmiller M, Braun H. "A direct adaptive method of faster backpropagation learning: The rprop algorithm. In *IEEE International Conference on Neural Networks*, San Francisco, USA; **1993**:586–591.

[23] Chan TSN. Butler matrix feed configuration for phased array. [Undergraduate thesis]. University of Queensland, Department of Electrical and Computer Engineering, Australia; **1994**.

AUTHOR BIOGRAPHIES

YOUSSEF HARKOUSS received the PhD degree in electronics from the University of Limoges, France, in 1998. From 1999 to 2000, he was Research Engineer at the CNRS, France. He joined the Lebanese University-Faculty of Engineering III in 2001 as a professor. His research interests include neural networks, fuzzy logic, fuzzy neural networks software development, RF and microwave modeling and optimization techniques, CAD of passive devices, smart antennas, and intelligent systems.

HASSAN SHRAIM received the PhD degree in industrial automation and control from the University of AIX Marseille III, France, in 2007. He joined the Lebanese University-Faculty of Engineering III in 2011 as an associate professor. His research interests include nonlinear control, fault diagnosis and fault tolerant control, neural networks, fuzzy logic, and intelligent systems.

HUSSEIN BAZZI received the PhD degree in mechanical engineering from the University of Sherbrooke, Canada, in 1998. He joined the Lebanese University, Faculty of Engineering III in 1999 as an associate professor. His research interests include numerical simulation in fluid mechanics and heat transfer, neural networks, fuzzy logic, and intelligent systems.

How to cite this article: Harkouss Y, Shraim H, Bazzi H. Direction of arrival estimation for smart antenna in multipath environment using convolutional neural network. *Int J RF Microw Comput Aided Eng*. 2018;28: e21282. https://doi.org/10.1002/mmce.21282