CSI-based Fingerprinting for Indoor Localization with Multi-scale Convolutional Neural Network

2021 IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE) | 978-1-6654-4516-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ECICE52819.2021.9645641

Zhuqi Shi
National Engineering Research Center of Communications & Networking, Nanjing University of Posts and Telecommunications
Nanjing 210003, China

Li Wei*
National Engineering Research Center of Communications & Networking, Nanjing University of Posts and Telecommunications
Nanjing 210003, China
weili@njupt.edu.cn

Youyun Xu
National Engineering Research Center of Communications & Networking, Nanjing University of Posts and Telecommunications
Nanjing 210003, China
yyxu@njupt.edu.cn

***Abstract*—With the rapid-growing demand for location-based services in indoor environments, fingerprint-based indoor positioning methods have attracted great interest owing to high accuracy and low online complexity. In this paper, we use the channel state information (CSI) of the massive MIMO (MaMIMO) system as the fingerprint to construct the fingerprint database. Different from the previous methods that only use CSI amplitude to construct fingerprints, phase information and angle of arrival (AOA) are added to the fingerprint to enhance the characteristics of fingerprint data. We modified the network according to the characteristics of fingerprint data based on Google Net and implemented a GoogleNet-like convolutional neural network(CNN) which uses skip connection and 1-D convolution kernel for fingerprint positioning. Experiment results show that with sufficient representative data sets, centimeter-level positioning can be achieved by using the proposed neural network, and the positioning accuracy can be further improved by 10% with the use of AOA information.**

Keywords—fingerprint location, CSI, MaMIMO, CNN, AOA

I. INTRODUCTION

Massive MIMO is an emerging technology used in 5G communication networks, which can greatly improve the spectrum efficiency of wireless systems by combining a large number of base station (BS) antennas with signal processing based on measured channel state information (CSI). CSI is, then, estimated by using pilot sequences during uplink transmission. The combination of a large number of antennas and accurate CSI enables the BS to multiplex users in the spatial domain. The theory shows that as the number of base station antennas increases, the performance of the system is only limited by the accuracy of the channel state information. Therefore, it makes sense to use a large antenna array to oversample the spatial information of the wireless channel.

Fingerprint recognition is a low-latency wireless positioning solution suitable for challenging environments such as indoors. The fingerprint method realizes positioning through the construction of the offline fingerprint database and online fingerprint matching. In the offline phase, appropriate position reference points are selected according to the layout of the application scenario and fingerprint information on each reference point in the online positioning phase, according to a specific matching algorithm (such as KNN, K-Means, and so on) The fingerprint data is matched with the existing data in the database, and the location of the point to be measured is estimated.

Received signal strength (RSS) is often used for indoor positioning due to its simplicity and low hardware requirements. The Radar system uses the received RSS to estimate the relative distance for triangulation [1]. However, RSS-based systems have two major shortcomings. Firstly, due to the multipath effect in the indoor environment, RSS varies greatly over time; secondly, RSS is coarse-grained signal information and cannot obtain richer multipath information. Previous fingerprint positioning schemes often used RSS as fingerprint information, but due to the inherent disadvantages of RSS, the accuracy of the fingerprint positioning method based on RSS is low.

Compared with RSS, CSI is a fine-grained physical layer (PHY) information that can provide detailed channel information at the subcarrier level [2]. Studies have shown that the accuracy of indoor location estimation based on CSI is significantly higher than that of RSS. Kotaru *et al.* used super-resolution algorithms and filtering-estimation techniques in their SpotFi algorithm to obtain high-precision AOA estimation from CSI information and achieved centimeter-level positioning accuracy [3]. Chen *et al.* adopt autoregressive (AR) modeling entropy of CSI amplitude as location fingerprint, and integrated AOA to enrich the fingerprint information [4], achieved good results, but the AR method is too rough, AOA is not enough to fully represent the phase information of CSI.

Deep learning (DL) has been widely used in the field of indoor positioning due to its powerful modeling capabilities and data feature extraction capabilities of complex models. Keun *et al.* applied deep learning to the training of the RSS fingerprint database and proved that its positioning effect is better than the non-deep learning KNN algorithm [5]. Vieira *et al.* used a conversion function to convert CSI information into a sparse form and used CNN to learn the features of the form to obtain decimeter-level positioning [6]. Jing *et al.* proposed a dual-stream 3D convolutional neural network, which uses two parallel subnets to learn the amplitude and phase information of CSI at the same time, and fused the probabilistic classification results in the final output layer to achieve the indoor positioning requirements [7].

This paper mainly studies an indoor CSI fingerprint location method based on deep learning. The collected CSI in the frequency domain and time domain is decomposed to obtain the signal characteristics, and use the Multiple Signal Classification (MUSIC) algorithm [8] to obtain the Angle of Arrival (AOA) to assist in the generation of fingerprint information, finally, a GoogleNet-like convolutional neural network is used for learning and training. The main contributions of this paper are as follows.

(1) Provide a fingerprint creation idea, make full use of the amplitude and phase information in the CSI, and add the AOA feature in the traditional positioning to the fingerprint.

(2) A convolutional neural network is designed, which has the following characteristics.

ISBN: 978-1-6654-4516-0

- Drawing lessons from the design ideas of the Inception network, the network performance is improved by increasing the width.
- Using skip connection to make up for the gradient loss caused by the network being too deep.
- Based on the characteristics of CSI fingerprint data, 1-D convolution kernel is used in the network.

II. CHANNEL FINGERPRINTING AND PREPROCESSING

A. Collection of Data Sets

The data used in this paper is obtained from KU Leuven Massive MIMO, which is a high-density labeled CSI fingerprint data with an error of less than 1 mm.

The base station (BS) deploys 64 antennas, an 8 × 8 uniform rectangular antenna array (URA), the distance between adjacent antenna elements is 70 mm, and the antenna array is placed 1 m away from the workbench (Red area in Fig. 1) which sends or receives signal at the same time. These 64 antennas are used to receive pilot signals from users and estimate CSI based on these pilot signals. The user uses a signal with a center frequency (fc) of 2.61 GHz and a bandwidth of 20 MHz. The pilot signal is composed of 100 subcarriers, and the frequency is evenly distributed. During the measurement process, a single-antenna user was placed indoors in an office and moved equidistantly along a predetermined path at 5mm intervals. The path was carried out along a grid spanning an area of 1.25 x 1.25 m. All transmission and reception routes were line-of-sight (los) and a total of 252004 sets of data were collected.

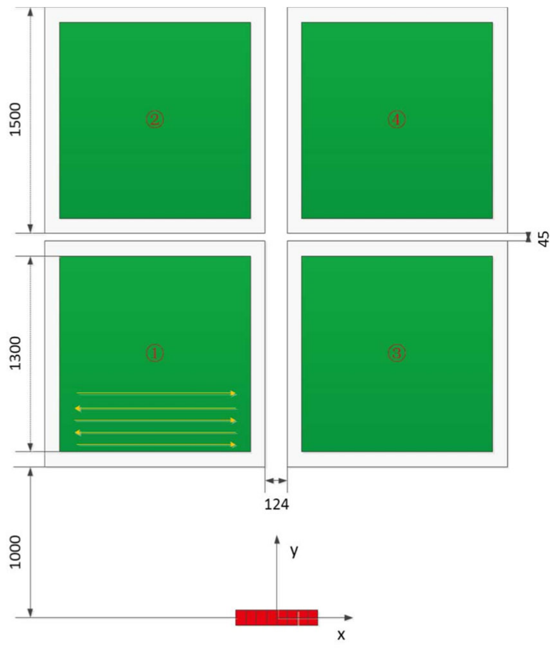

Fig. 1. Information measurement scenario, using URA antenna array. The user moves to the green area. All measurements on the figure are in millimeters.

B. Data Preprocessing

After preprocessing the signal received in the previous section such as interpolation, filtering, and denoising, the standard CSI matrix is obtained as shown below.

$$CSI_{matrix} = \begin{bmatrix} csi_{1,1} & \cdots & csi_{1,n} \\ \vdots & \ddots & \vdots \\ csi_{m,1} & \cdots & csi_{m,n} \end{bmatrix} \quad (1)$$

where $csi_{m,n}$ represents the CSI value of the m^{th} antenna and the n^{th} subcarrier. In this paper, m=64, and n=100.

It is well known that multipath propagation of the signal is manifested as delay extension in a time domain and selective fading in a frequency domain. The CSI data we collected is a complex value. It is converted from the complex domain to the extreme domain, which is the channel frequency response (CFR). It uses amplitude-frequency characteristics and phase-frequency characteristics to describe the multipath propagation of signals, as shown below.

$$H(k) = \|H(k)\| e^{j\angle H(k)} \quad (2)$$

where $H(k)$ is the CSI of the k^{th} subcarrier, $\|H(k)\|$ is the amplitude response of the k^{th} subcarrier, and $\angle H(k)$ is the phase response of the k^{th} subcarrier.

IFFT on CFR is performed to obtain the channel impulse response CIR of the CSI in the time domain, as shown below.

$$h(\tau) = \sum_{i=1}^{L} a_i e^{-i\theta_i} \delta(\tau - \tau_i) \quad (3)$$

where a_i is the amplitude attenuation of the i^{th} path, θ_i is the phase offset of the i^{th} path, τ_i is the time delay of the i^{th} path, L is the number of multipaths, and $\delta(\tau)$ is the Dirac impulse function.

By concatenating the three sets of features (raw CSI, CFR, CIR) above, we can get the preprocessed initial input CSI matrix $R^{64\times100\times6}$,as show in Fig. 2.

C. MUSIC Algorithm to Estimate AOA

The fingerprint location algorithm compares the measured data with the fingerprint database data and selects a number of the most similar data to calculate the estimated coordinates based on the weighted average of similarity. Therefore, there may be similar fingerprints in adjacent locations that may lead to misjudgment, so we add the spatial characteristic AOA to help the system determine the approximate position through AOA and eliminate the interference of some similar items. At present, the most common method to obtain the AOA value is to use the Multiple Signal Classification(MUSIC) algorithms.

The MUSIC algorithm is based on the subspace decomposition of the far-field narrowband signal. It uses the orthogonality of the signal subspace and the noise subspace to construct a spatial spectrum function and estimates the signal parameters through spectral peak search.

In reality, due to the influence of multipath, the spatial spectrum has multiple peaks, and we choose the angle corresponding to the highest peak as the AOA of the direct path.

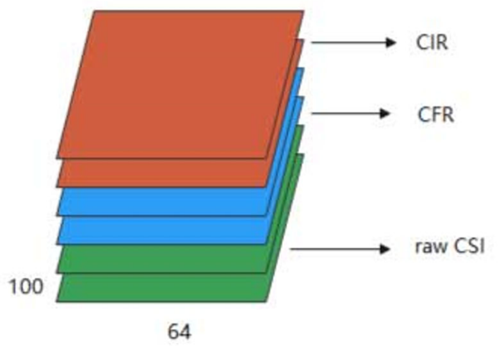

Fig. 2. structure of CSI-input-matrix,They are the real and imaginary parts of the original CSI,the amplitude response and phase response of CFR,and the amplitude attenuation and phase shift of CIR.

III.MODEL OF CONVOLUTIONAL NETWORK

Convolutional Neural Networks (CNN) use trainable filters to convolve data to extract features and information, being effective in extracting structured features from complex data. The collected MaMIMO CSI contains a large amount of structured data. Therefore, it is feasible to apply CNN to spatial information inference [9]. This section first introduces the specific structure of the CNN and then discusses settings for the internal parameters of the CNN.

A. Structure of CNN

In order to infer the user's spatial position from the MaMIMO CSI with a large amount of data, the shallow neural network no longer completes this job effectively, so it becomes necessary to design a deep neural network. Generally speaking, the most direct way to improve network performance is to increase the depth and width of the network. However, this also brings two problems: one is the rapid increase in the number of network parameters that causes the network to fall into overfitting, and the other consumes huge computing resources.

We draw on the design concept of Google net (also known as Inception module) [10] which converts full connection or even convolution into the sparse connection as the actual biological neural network connection itself is sparse. Besides, Arora *et al*. proved that for a large-scale sparse network, the optimal network topology is constructed layer after layer by analyzing the correlation statistics of the preceding layer activations and clustering neurons with highly correlated outputs [11]. This method reduces the number of network parameters without loss of performance.

We designed a GoogleNet-like convolutional neural network, which greatly reduced the number of Inception modules, and replaced the softmax prediction layer with jump connections. The network structure is shown in Fig. 3.

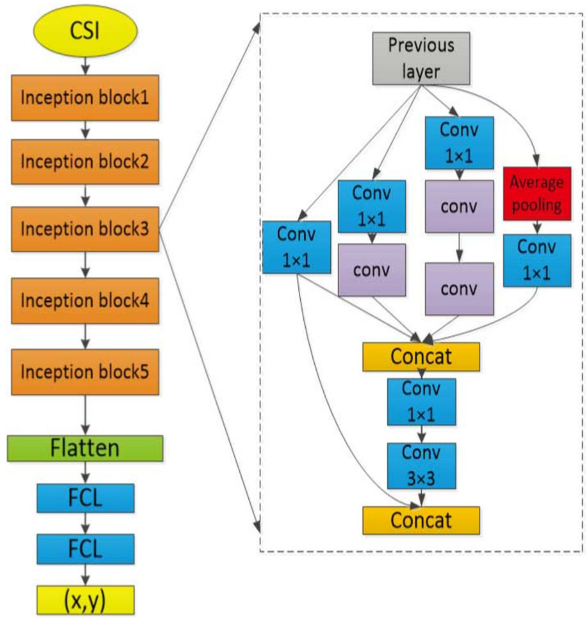

Fig. 3. Architecture of the network, see section B for details of the convolution kernel parameters

The network consists of two parts in total. The first part is composed of 5 convolution modules (also called inception blocks), and the second part is the fully connected layer. After passing through the two fully connected layers, the position information is finally obtained.

The entire CNN contains 15-convolutional layers, with a trainable parameter of 129818, which is far lower than the number of parameters used by the network with the same depth, which benefits from the use of a large number of 1×1 convolutions (also known as Network in Network)[12]. 1×1 convolution has two functions: firstly, they are mainly used as a dimensionality reduction module to eliminate computational bottlenecks to increase the size of the network; secondly, It can effectively reduce the number of convolution parameters under the premise that the receptive field of the convolution kernel remains unchanged.

The branch of the Inception block uses three convolutional layers of different sizes. The two consecutive convolutional layers of the third branch are equivalent to a larger convolutional layer. This is also done to minimize the parameters under the premise of the same receptive field [13]. The use of different sizes of convolution kernels for each branch means different sizes of receptive fields, and the final stitching means the fusion of features of different scales, which facilitates us to obtain richer feature information.

A network that is too deep may have the problem of vanishing gradients. To solve this problem, skip connections are used [14]. We connect the output of the first branch of the inception block to the total output as the input of the next inception block. Such a structure can compensate for the gradient loss caused by stepwise propagation in the depth structure, which is beneficial to the error propagation of the depth structure.

The network also uses Batch Normalization(BN) [15]. BN overcomes the shortcomings of traditional neural networks that use a small learning rate and replace it with a larger one, greatly reducing the number of iterations required to achieve the original accuracy rate, and greatly shortening the training time. As BN plays the role of regularization, we can reduce or cancel Dropout and simplify the network structure.

TABLE I. KERNEL_SIZE AND CONVOLUTION STRIDE USED IN INCEPTION MODULE, THE AVERAGE POOLING SIZE IS THE SAME AS THE KERNEL_SIZE

CONV	Inception block 1	Inception block 2	Inception block 3	Inception block 4	Inception block 5
kernel_size	(1,9)	(1,9)	(1,9)	(9,1)	(9,1)
stride	(1,2)	(1,2)	(1,5)	(2,1)	(2,1)

In view of the difference between MaMIMO CSI and ordinary visual pictures, the main data characteristics of CSI no longer is distributed like pictures but have their distribution in the two dimensions of subcarriers and antennas, so 2-D convolution kernels (such as 3 × 3 and 5 × 5) is not suitable for the network in this situation, so we use 1-D convolution kernels instead. Our method is to first perform 1-D convolution on one dimension (subcarrier) of CSI, and then perform convolution on another(antenna). The specific operation is to use the (1,9) convolution kernel in the first three Inception blocks to extract the feature value in the subcarrier dimension, and use the (9,1) convolution kernel to extract the feature value in the antenna dimension in the remaining two Inception blocks. Experiments have proved that this helps to extract signal features better. The parameters used in the network are shown in Table I.

IV. EXPERIMENT EVALUATION

In this section, we discuss the factors that affect the CNN positioning performance. First, the performance differences of CNN are analyzed by using different convolution kernels and then compare it with other positioning schemes in the same data set. All training uses 85% of the data as the training set, 10% as the validation set, and 5% as the test set.

In this paper, the root mean square error (RMSE) is used as the loss criterion, and the RMSE calculation is as follows.

$$RMSE = \frac{\sum\sqrt{(p-\hat{p})^2}}{n} \tag{4}$$

where p is the user's actual measurement position (x, y), $\hat{p}$ is the user's predicted position $(\hat{x}, \hat{y})$, all expressed in millimeters, and n is the total number of sampling points.

A. Influence of adding AOA

The purpose of using AOA is to distinguish difficult data through spatial information and to narrow the scope of positioning. We multiply the AOA value and the all-one matrix of (64,100,2) and merge it into the initial input matrix I, that is, the initial input matrix becomes $I' = R^{64\times100\times8}$. In order to avoid data skew, it is necessary to normalize the AOA matrix and the CSI matrix.

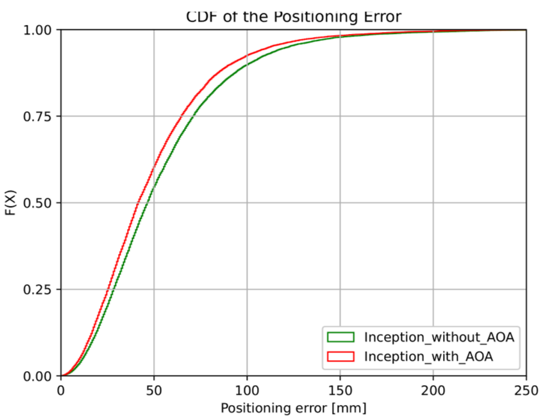

Fig. 4. CDF of positioning error for fingerprint with or without AOA

Figure 4 shows the impact of adding AOA to fingerprints or not. The addition of AOA further improves the positioning performance of the system, which proves the correctness of our theory.

Table II shows that the AOA scheme has achieved a mean square error of 49mm, and the positioning accuracy is improved by 10% compared to the method without AOA. The median error increased by 12%.

TABLE II. POSITIONING ERROR FOR FINGERPRINT WITH OR WITHOUT AOA

	RMSE[mm]	Median error[mm]
Incetion-with-AOA	49	41
Incetion-without-AOA	54.6	56.6

Although the positioning accuracy of the latter is improved compared with the fingerprint construction method without AOA, the improvement is not obvious. Perhaps because the AOA information obtained in the article is calculated based on the most basic MUSIC algorithm, we have not improved and optimized it. There are many errors in the AOA value that affect the final positioning effect.

B. Compare with other methods

MaMIMO CSI has rich information and requires a certain depth of neural network to be effectively trained, it is unfair to compare the network in this paper with the shallow neural network, so we have implemented a 15-layer ResNet [16]. In addition, we also implemented three other positioning schemes, namely a neural network with three fully connected layers [5], SpotFi algorithm [3], and KNN algorithm [17] to compare performance.

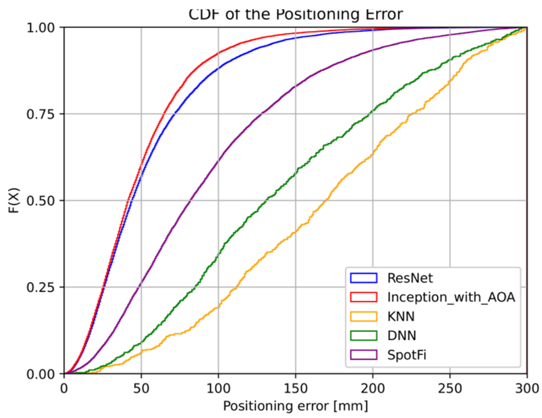

Fig. 5. CDF of positioning error for different positioning schemes

Figure 5 shows the positioning accuracy of several different schemes. Inception Net and ResNet have achieved performance far exceeding the other schemes by their deep network and excellent design. The specific parameters are shown in Table III. The mean square error and median error of InceptionNet are lower than ResNet, and the amount of network parameters is reduced by 40% compared with ResNet, which greatly reduces the complexity of network operation and saves computing resources.

TABLE III. POSITIONING ERROR AND TRAINABLE PARAMETERS OF DIFFERENT SCHEMES.

	RMSE[mm]	Median error [mm]	Number of trainable parameters
Inception	49	41	129818
ResNet	55.4	43.3	217378
SpotFi	99	82.5	None
DNN	190	132.5	3860502
KNN	232.5	172.5	None

V. CONCLUSION

Massive MIMO is an emerging technology that is being widely used. We have studied the feasibility of positioning convolutional neural networks in MIMO systems and designed a type of GoogleNet-like convolutional neural network based on the measured MaMIMO CSI. According to the characteristics of CSI data, a more suitable convolution kernel was selected, and average positioning accuracy of 54.6 mm was achieved. We also calculated the AOA by the MUSIC algorithm, and incorporated it into the fingerprints, proving that the addition of AOA can assist positioning and improve the positioning performance of the network, further increasing the accuracy by 10.2%, and finally reaching the average accuracy of 49 mm.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation Incubation project of Nanjing University of Posts and Telecommunications (NY220009).

REFERENCES

[1] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," Proceedings IEEE INFOCOM 2000. pp. 775-784 vol.2.

[2] Z. Yang, Z. Zhou and Y. Liu , "From RSSI to CSI: Indoor Localization via Channel Response," Acm Computing Surveys, 2013, 46(2):1-32.

[3] M. Kotaru, K. Joshi, D. Bharadia and K. Sachin, "SpotFi: Decimeter Level Localization Using WiFi," Acm Sigcomm Computer Communication Review, 2015, 45(4):269-282.

[4] L. Chen, I. Ahriz and D. Le Ruyet, "AoA-Aware Probabilistic Indoor Location Fingerprinting Using Channel State Information," IEEE Internet of Things Journal,doi:10.1109/JIOT.2020.2990314.

[5] K. Y. Seok and J. H. Lee, "Deep Learning Based Fingerprinting Scheme for Wireless Positioning," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 312-314

[6] J. Vieira, E. Leitinger, M. Sarajlic, X. Li and F. Tufvesson, "Deep convolutional neural networks for massive MIMO fingerprint-based positioning," 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017, pp. 1-6

[7] Y. Jing, J. Hao and P. Li, "Learning Spatiotemporal Features of CSI for Indoor Localization With Dual-Stream 3D Convolutional Neural Networks," in IEEE Access, vol. 7, pp. 147571-147585, 2019, doi: 10.1109/ACCESS.2019.2946870.

[8] R. Schmidt, "Multiple emitter location and signal parameter estimation," in IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276-280, March 1986.

[9] M. Arnold, S. Dorner, S. Cammerer and S. Ten Brink, "On Deep Learning-Based Massive MIMO Indoor User Localization," 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Kalamata. 2018.

[10] C. Szegedy,W. Liu, Y. Jia1, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9.

[11] S. Arora, A. Bhaskara, R. Ge and T. Ma, "Provable Bounds for Learning Some Deep Representations," Computer Science, 2013.

[12] M. Lin, Q. Chen, and S. Yan, "Network in network," CoRR, abs/1312.4400, 2013.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826.

[14] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017.

[15] S. loffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," JMLR.org, 2015.

[16] S. D. Bast, A. P. Guevara and S. Pollin, "CSI-based Positioning in Massive MIMO systems using Convolutional Neural Networks," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 2020, pp. 1-5.

[17] Q. Song, S. Guo, X. Liu and Y. Yang, "CSI Amplitude Fingerprinting-Based NB-IoT Indoor Localization," in IEEE Internet of Things Journal, vol. 5, no. 3, pp. 1494-1504, June 2018