Multi-Sources Information Fusion Learning for Multi-Points NLOS Localization

Bohao Wang*, Fenghao Zhu*, Mengbing Liu§, Chongwen Huang*, Qianqian Yang*,
Ahmed Alhammadi‡, Zhaoyang Zhang*, and Mérouane Debbah¶‖, *Fellow, IEEE*

* College of Information Science and Electronic Engineering, Zhejiang University, 310027, Hangzhou, China
‡ Technology Innovation Institute, 9639 Masdar City, Abu Dhabi, UAE
§ School of Electrical and Electronics Engineering, Nanyang Technological University, 639798, Singapore
¶ KU 6G Research Center, Khalifa University of Science and Technology, P O Box 127788, Abu Dhabi, UAE
‖ CentraleSupelec, University Paris-Saclay, 91192 Gif-sur-Yvette, France

Abstract—Accurate localization of mobile terminals is crucial for integrated sensing and communication systems. Existing fingerprint localization methods, which deduce coordinates from channel information in pre-defined rectangular areas, struggle with the heterogeneous fingerprint distribution inherent in non-line-of-sight (NLOS) scenarios. To address the problem, we introduce a novel multi-source information fusion learning framework referred to as the Autosync Multi-Domain NLOS Localization (AMDNLoc). Specifically, AMDNLoc employs a two-stage matched filter fused with a target tracking algorithm and iterative centroid-based clustering to automatically and irregularly segment NLOS regions, ensuring uniform fingerprint distribution within channel state information across frequency, power, and time-delay domains. Additionally, the framework utilizes a segment-specific linear classifier array, coupled with deep residual network-based feature extraction and fusion, to establish the correlation function between fingerprint features and coordinates within these regions. Simulation results demonstrate that AMDNLoc significantly enhances localization accuracy by over 55% compared with traditional convolutional neural network on the wireless artificial intelligence research dataset.

Index Terms—Multi-sources, information fusion, fingerprint localization, inverse, heterogeneity, regional covariant.

I. INTRODUCTION

As a key usage scenario for sixth-generation communication, integrated sensing and communication requires a low-latency and high-precision mobile terminal (MT) localization, especially in fields like the smart city, internet of vehicles, telemedicine and so on. Among various methods, fingerprint-based localization, leveraging unique multi-path features at each location to infer specific positions, stands out in massive multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) systems.

In particular, it treats channel state information (CSI) of multi-path features as fingerprints, and matches real-time measurements with stored fingerprints to estimate locations. However, this process, especially in outdoor multi-point non-line-of-sight (NLOS) scenarios, is fraught with complexities. The multi-path features that are influenced by diverse urban elements like buildings and scattered objects lead to substantial heterogeneity in fingerprint distribution. This results in an uneven regional on the relationship between fingerprints

and locations, complicating the inverse problem of location inference from fingerprints [1].

To tackle the issue, traditional solutions involve dividing a large area into grid cells, assigning labels within each cell, and iteratively refining the search until the closest fingerprint in the dataset is matched [2–10]. Although manual division of areas is easy to operate, the localization accuracy can be sensitive to the specifics of regional divisions. Furthermore, due to the heterogeneity issue, while fingerprints of adjacent positions are more likely to exhibit high similarity, existing works do not guarantee that all fingerprints within a cell share a high similarity, no matter the cell is regular or not.

Building upon the manually divided areas, several artificial intelligent (AI) techniques have been proposed to build the inverse correlation function between fingerprints and locations [5–11]. However, these approaches refine the search area hierarchically which means that there are amounts of parameters to be updated and once the coarse position of the fingerprint is predicted incorrectly, errors will accumulate to the final results. What's more, though each location's multi-path features are unique, these approaches may fail to fully encapsulate this uniqueness by only choosing one of the received signal strength (RSS), channel impulse response (CIR), channel frequency response (CFR), and the angle-delay channel amplitude matrix (ADCAM) as the fingerprint. Experiments suggest that there exists a proportion of remote points sharing very similar fingerprints within a scene in different dataset, thereby violating the foundational assumption of independent and identically distributed (i.i.d) data and undermining the predicted results [12, 13].

To solve above-mentioned issues, we propose a novel multi-source information fusion learning framework for large scale outdoor multi-points NLOS localization called the autosync multi-domain NLOS localization (AMDNLoc) framework, which automatically and irregularly generates the classification regions based on the fusion of frequency, power and time-delay information of CSI for the first time. Specifically, we develop a two-stage matched filter to identify and extract all parallel feature intervals related to CFR distribution, termed parallel feature CFR (PFCFR), and classify all CFR images
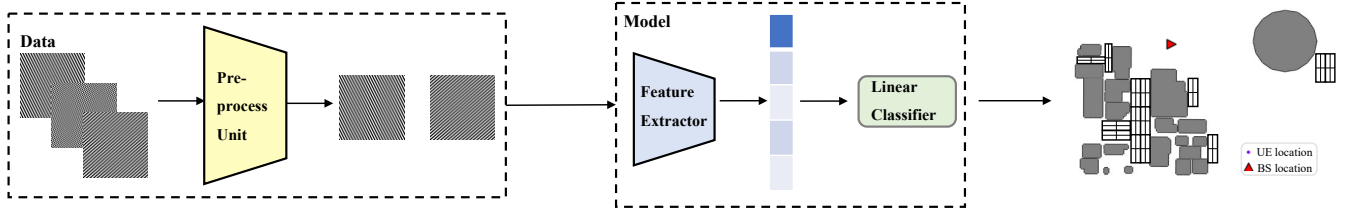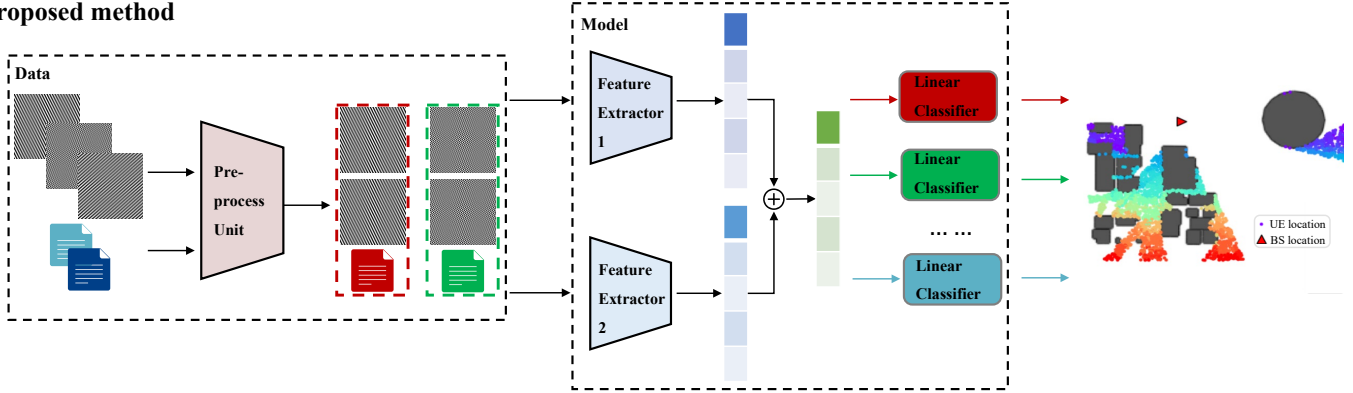
Figure 1: System Diagram of proposed AMDNLoc framework compared with existing methods.

according to the closest PFCFR by target tracking algorithm. Then, we utilize the iterative centroid-based clustering to partition the ADCAM considering power, angle, and delay characteristics, and combine these two classification areas maintaining the same fingerprint distribution characteristics. Then, special samples are eliminated by data cleaning and a segment-specific linear classifier array coupled with deep residual network-based feature extraction and fusion is introduced to inversely map the fingerprints to the coordinates in the respective regions while maintaining the global features. State of the art (SOTA) results are achieved in the ray-tracing wireless AI research dataset (WAIR-D) [14, 15]. Also the visualization of classification areas and comparison of multiple processing are validated to ensure the enhanced adaptability and scalability of AMDNLoc. The code is avalable in the [16].

II. SYSTEM AND CHANNEL MODEL

In this section, we begin by presenting the channel model and subsequently formulate the heterogeneity problem in NLOS fingerprint localization.

The channel transmission diagram in a MIMO-OFDM system is illustrated in Fig. 2. The BS is equipped with a uniform linear array (ULA) comprising N_t antennas. We have $m \in \{1, ..., M\}$ MTs, each equipped with a single omni-directional antenna. The angle of arrival (AOA) and the physical distance between the transmit antenna and the first receive antenna associated with the p-th path are denoted by $\phi_{p,m} \in (0, \pi)$ and $d_{p,m}$, respectively.

The CIR vector associated with the p-th path of the m-th user is given by:

$$\mathbf{q}_{p,m} = a_{p,m} \mathbf{e}(\phi_{p,m}), \tag{1}$$

where $a_{p,m} \sim \mathcal{CN}(0, \sigma_{p,m})$ represents the complex gain associated with the p-th path, and $\mathbf{e}(\phi)$ is the array response vector corresponding to the AOA ϕ. It takes the form:

$$\mathbf{e}(\phi) = \left[1, e^{-\jmath 2\pi \frac{d \cos(\phi)}{\lambda_c}}, \ldots, e^{-\jmath 2\pi \frac{(N_t-1) d \cos(\phi)}{\lambda_c}} \right]^T, \tag{2}$$

where $j = \sqrt{-1}$, d is the antenna spacing (typically $\lambda/2$ in recent MIMO systems); $\tau_{p,m} = n_{p,m} T_s$ represents the distinguishable propagation delay associated with the p-th path, $n_{p,m}$ refers to the sampled delay for the p-th path, and T_s denotes the sample interval. The CFR is expressed as a sum of time-domain CIRs with varying delays:

$$\mathbf{h}_{m,l} = \sum_{p=1}^{P} a_{p,m} \mathbf{e}(\phi_{p,m}) e^{-\jmath 2\pi \frac{l n_{p,m}}{N_c}}, \tag{3}$$

where N_c is the number of subcarriers in the OFDM system, $l \in \{1, ..., N_c\}$. Then, the overall CFR matrix known to the BS can be denoted as the stack of $\mathbf{h}_{m,l}$, i.e.,

$$\mathbf{H}_m = [\mathbf{h}_{m,0}, \mathbf{h}_{m,1}, \ldots, \mathbf{h}_{m,N_c-1}], \tag{4}$$

Then, authors in [4] established a mapping from the CFR matrix to a sparse structure by employing DFT operations, referred to as the angle-delay channel response matrix. It is a linear transformation of the CSI computed by multiplying it with two DFT matrices.

Let us define the DFT matrix $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$ as:

$$[\mathbf{V}]_{z,q} \triangleq \frac{1}{\sqrt{N_t}} e^{-j 2\pi \frac{\left(z \left(q - \frac{N_t}{2} \right) \right)}{N_t}},$$

and $\mathbf{F} \in \mathbb{C}^{N_c \times N_c}$ as:

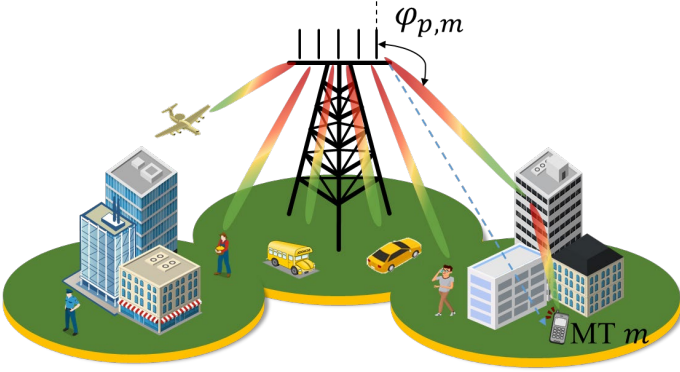$$[\mathbf{F}]_{z,q} \triangleq \frac{1}{\sqrt{N_c}} e^{-j 2\pi \frac{zq}{N_c}},$$

Figure 2: The wireless channel from MT m to the BS.

Therefore, the ADCAM of user m is defined as:

$$[\mathbf{A_m}]_{z,q} = \mathbb{E}\left\{\left|\left[\mathbf{V}^H\mathbf{H}_m\mathbf{F}\right]_{z,q}\right|\right\}, \tag{5}$$

where the (z,q) element of the ADCAM represents the absolute gain of the z-th AOA and q-th delay of the channel.

As we know, fingerprint is determined by the locations and a specific configuration of the propagation environment, given by:

$$Fingerprint = \mathcal{F}(\boldsymbol{p}, pv), \tag{6}$$

where *Fingerprint* represents the channel characteristics used for localization. $\boldsymbol{p} := [x, y]$ denotes the 2D positional coordinates of the users, and pv encompasses the system parameters which remain constant during training and testing.

Traditional fingerprint localization aims to inversely map the *Fingerprint* back to $\boldsymbol{p}$. However, the impact of different buildings on various areas differs, so that changes in the fingerprint in different areas play different roles in affecting user locations. This leads to the issue of fingerprint heterogeneity in distribution. Therefore, the inverse process of fingerprint extraction should be expressed as:

$$\hat{\boldsymbol{p}}_{\text{real}} = \mathcal{F}^{-1}(Fingerprint, pv, cv), \tag{7}$$

where cv represents a covariate related to the region, weather, and other factors. Our objective in this research is to eliminate the regional covariate from fingerprint localization and establish a one-to-one relationship solely between $\boldsymbol{p}$ and *Fingerprint*.

III. AMDNLOC MULTI-SOURCES FRAMEWORK

In this section, we delve into the heterogeneity matched filter, the design of the network structure, and the detailed network training methodology within the AMDNLoc framework.

A. Heterogeneity Matched Filter

In tackling the complex issue of heterogeneous fingerprint distribution, our research has led to the development of a specialized classification approach to uniquely addresses the distinct distribution characteristics inherent to both CFR and ADCAM.
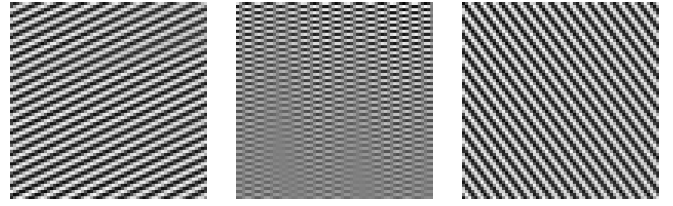

(a) (101.22 49.29) (b) (73.85 30.77) (c) (144.64 93.75)

Figure 3: CFR example figure of randomly selected MTs in the 00743 scenario of WAIR-D

1) PFCFR: We commence by depicting the matrix H from Eq. 4 as a two-channel grayscale image, as illustrated in Fig. 3. In this representation, the x-axis corresponds to the carrier frequency, while the y-axis denotes the antenna pair. A consistent observation across the dataset reveals the presence of distinct horizontal and vertical translational structures within certain regions in CFR, which we refer to as PFCFR. PFCFR is a key indicator of data distribution of CFR. This distribution is largely shaped by physical phenomena such as refraction, reflection, and diffraction, which exhibit more pronounced similarities in adjacent positions. Consequently, it is a logical inference that CFR sharing similar PFCFR characteristics are likely to be located in geographically close areas. This proximity-based similarity allows for an effective categorization of these CFR according to their respective PFCFR values, providing valuable insights into the spatial dynamics of data distribution influenced by channel properties.

To leverage this insight, we propose a novel two-stage pipeline matched filter outlined in Alg. 1 and Alg. 2. Inspired by the target tracking algorithm, the filter is designed firstly to extract PFCFR from all samples as the template and CFR as the source image for CFR categorization. To mitigate the risk of coincidental similarities in specific regions between two images, our approach incorporates dual template regions. We define $T_1(i, j, x', y')$ and $T_2(i, j, x', y')$ to represent the pixel values from the upper-left and lower-right corners, respectively, of the j-th source image. Here, the notation (i, x', y') specifies the pixel coordinates in the i-th template image. The application of this filter involves systematically shifting the origin of the template image across each point in the source image. The similarity between the template and source images is calculated by aggregating the products of their corresponding pixel values across the entire span of the template. To quantify this similarity, we employ the normalized cross-correlation, denoted as $\mathbf{E_n(i,j)}$, given by:

$$\mathbf{E_n(i,j)} = \max \frac{\sum_{x',y'} T_n(i,j,x',y') I(j, x+x', y+y')}{\sqrt{\sum_{x',y'} T_n(i,j,x',y')^2 \sum_{x',y'} I(j, x+x', y+y')^2}}, \tag{8}$$

where $n = 1, 2$, $\mathbf{E_n(i,j)}$ is between 0 and 1, and the closer it is to 1, the higher the similarity. Consequently, the process of our proposed two-stage pipeline matched filter can be summarized as follows:

Initialization: We set τ_c to a value greater than M, which serves as a threshold for determining whether the j-th image is matched. Then, τ_{in} and τ_{out} are set as parameters indicative of

spatial proximity within and between categories, respectively. Additionally, we use $class\_num$ to denote the number of categories.

Match Within Categories: In this stage, we initiate by organizing the sequence of all images into a list denoted as $path\_List$. Our process starts with the first image in this list, for which we select two distinct templates, represented as $T_n(1, j, x', y')$, where $n = (1, 2)$. The subsequent step involves systematically matching the defined template area with corresponding areas in all subsequent images within the $path\_List$, employing Eq. 8 for this purpose. If the match exceeds τ_{in}, the two images are assigned to the same category. We then move on to the second image. If it hasn't been matched yet, we select two templates, $T_n(2, j, x', y')$, and repeat the matching process. We use the variable t to decide whether a template is matched or not during one matching process. It's important to note that if the template doesn't find any matching images, the match is sought in the preceding images along the path, starting from the first image. Given that these images have already been categorized, the corresponding category number of the matched image is assigned to the template. We then go to the next image, repeating the above steps. This process continues until each image has been assigned its own category number, denoted as c_j. The algorithm is shown as Alg. 1.

Match Between Categories: The objective of this stage is to macroscopically merge similar categories. We utilize Eq. 8 to evaluate the similarity between the templates derived from the first-stage classification results. When the similarity measurement between any two templates exceeds τ_{out}, we proceed to amalgamate all the CFR images associated with these templates into a single category. This merging process continues until there is no further change in the number of final classifications. The comprehensive algorithm for this stage is methodically outlined in Alg. 2.

2) ADCAM: To assist in the pre-classification of ADCAM, we employ the iterative centroid-based clustering method to partition information such as the AOD, AOA, gains, and pathloss of each path into K clusters based on their distances from K class centers. To choose the appropriate value of K, we combine the Silhouette coefficients and Calinski-Harabasz methods to combine both the direct distances and covariance [17]. This helps avoid scenarios where the graph of the clustering quality measure exhibits a smooth pattern (e.g., horizontal or continuously ascending/descending), making it difficult to ascertain the ideal K value.

3) Data Fusing and Cleansing: To ensure high-quality input data for our model, our initial step involves utilizing the two classification regions to refine the final classification categories. Subsequently, we apply data cleansing techniques to remove any anomalous fingerprints that might skew the analysis. In the phase of combining classifications, we adopt a systematic approach. For instance, if there are only Class 0 and Class 5 ADCAM images found within the Class 0 CFR category, we assign values of [0, 0] and [0, 5] to the new categories as Class 0 and 1, respectively. This reclassification

---

Algorithm 1 Match Within Categories

Input: All CFR images of M users, τ_{in}, τ_c, and the size of template a, b
Output: c_j for $j = 1, 2, ...M$
1: **Initialization:**
$\quad$ $class\_num = 0$ and set $c_j = \tau_c > M$ for
$\quad$ $j = 1, 2, ...M$
2: **for** $i = 1, 2, ..., M$ **do** $\qquad \triangleright$ Match within categories
3: $\quad$ $t \leftarrow 0$
4: $\quad$ **if** $i = 1$ or $c_i = \tau_c$ **then**
5: $\quad\quad$ Choose $T_n(i, j, x', y')$ according to a, b
6: $\quad\quad$ $c_i \leftarrow class\_num$
7: $\quad$ **end if**
8: $\quad$ **for** $j = i + 1, i + 2, ..., M$ **do**
9: $\quad\quad$ **if** $c_j = \tau_c$ **then**
10: $\quad\quad\quad$ Compute $E_n(i, j)$
11: $\quad\quad\quad$ **if** $E_n(i, j) >= \tau_{in}$ **then**
12: $\quad\quad\quad\quad$ $c_j \leftarrow class\_num, t \leftarrow 1$
13: $\quad\quad\quad$ **end if**
14: $\quad\quad$ **end if**
15: $\quad$ **end for**
16: $\quad$ **if** $t = 0$ **then**
17: $\quad\quad$ **for** $k = 1, 2, ..., i$ **do**
18: $\quad\quad\quad$ Compute $E_n(i, k)$
19: $\quad\quad\quad$ **if** $E_n(i, k) >= \tau_{in}$ **then**
20: $\quad\quad\quad\quad$ $t \leftarrow 1, c_j \leftarrow c_k$
21: $\quad\quad\quad\quad$ Break the loop
22: $\quad\quad\quad$ **end if**
23: $\quad\quad$ **end for**
24: $\quad$ **end if**
25: $\quad$ **if** $t = 0$ **then**
26: $\quad\quad$ $c_j \leftarrow class\_num$
27: $\quad$ **end if**
28: $\quad$ $class\_num \leftarrow class\_num + 1$
29: **end for**

Algorithm 2 Match Between Categories

Input: All templates and respective representative CFR images, τ_{out}, c_j and the size of template a, b
Output: c_j for $j = 1, 2, ...M$
1: Arrange c_j in ascending order and update the index of the arranged c_j to the new c_j for $j = 1, 2, ...M$
2: **for** $i = 1, 2, ..., max(c_j)$ **do** $\triangleright$ Match between categories
3: $\quad$ **for** $j = 1, 2, ..., max(c_j)$ **do**
4: $\quad\quad$ **if** $j \neq i$ **then**
5: $\quad\quad\quad$ Choose $T_n(i, j, x', y')$ according to a, b
6: $\quad\quad\quad$ Compute $E_n(i, j)$
7: $\quad\quad\quad$ **if** $E_n(i, k) >= \tau_{out}$ **then**
8: $\quad\quad\quad\quad$ Merge c_i and c_j into the same category
9: $\quad\quad\quad$ **end if**
10: $\quad\quad$ **end if**
11: $\quad$ **end for**
12: **end for**
13: Arrange c_j in ascending order and update the index of the arranged c_j to the new c_j for $j = 1, 2, ...M$
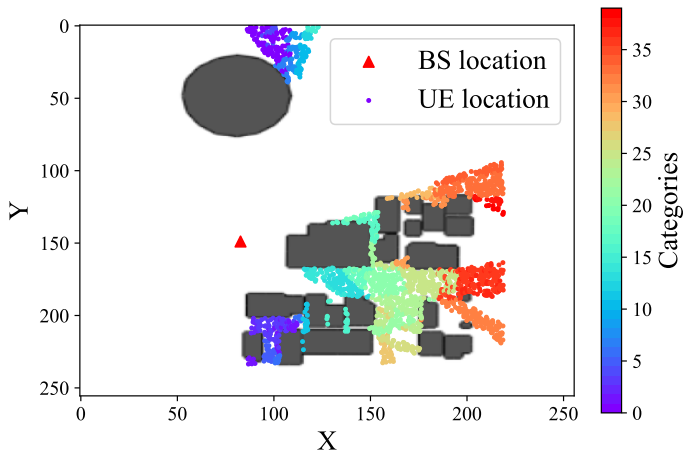
Figure 4: NLOS pre-classification.

process is methodically applied across all categories.

During the data cleansing phase, we eliminate any category that contains a sample size of 2 or fewer. The rationale behind this decision is that such a small sample size in a category typically indicates an aberration, possibly due to multi-path features that significantly diverges from neighboring points. This could be attributed to erroneous recordings or the point being situated in a distinctly unique location.

B. Network Training

After classifying the samples into different regions as shown in Fig. 4, we proceed with the network training process. This process consists of feature extraction, feature fusion, and filter regression, as illustrated in Fig. 1. The objective is to enable fingerprint localization, expressed as:

$$\hat{\mathbf{p}}_{i,j} = \mathcal{G}_{\mathbf{3,i}}(\mathcal{P}(\mathcal{G}_{\mathbf{1}}(H_{i,j}), \mathcal{G}_{\mathbf{2}}(A_{i,j}))), \quad (9)$$

where $\mathcal{G}_{\mathbf{1}}$ and $\mathcal{G}_{\mathbf{2}}$ represent the mappings built by deep residual network 18 (ResNet18 [18]) from CFR or ADCAM images to a vector with a flattened length of M_H or M_A respectively. $\mathcal{P}$ is a function that merges these two sets of features along the same dimensions. Finally, $\mathcal{G}_{\mathbf{3,i}}$ is a parallel multi-head mechanism in which the combined features are divided into different linear classifiers for regression operations. The variables $\sum_i i = I$ denote the number of final classifications, and each linear classifier corresponds to a specific region denoted as i. $\sum_i \sum_j i \times j = M$, and i, j represents the j-th sample in the i-th class of the fingerprint after the heterogeneity matching.

The loss function used for training is the mean-squared error (MSE) between the true positional coordinates p and the predicted positional coordinates $\hat{p}$, calculated as:

$$\text{MSE Loss} = \frac{1}{n} \sum_{i=1}^{n} \|\hat{\boldsymbol{p}}_i - \boldsymbol{p}_i\|^2, \quad (10)$$

Here, $n = 16$ indicates the mini-batch size. Weight initialization is performed using Xavier initialization, and the Adam optimizer is utilized. The initial learning rate is set to 0.003.

IV. NUMERICAL RESULTS

This section demonstrates our framework's performance through simulations on scenario 00743 of the WAIR-D dataset. Our approach was trained for 150 epochs with the Adam optimizer with a learning rate of 3×10^{-3} in the Nvidia 3090. Network architectures and hyperparameters were chosen via cross-validation for consistency across experiments unless stated otherwise.

A. Datasets

WAIR-D is a mmWave MIMO dataset for researchers to study and evaluate AI algorithms for wireless systems. It utilizes ray-tracing method, which is based on a high-frequency approximation to the Maxwell equations and describes the propagating field as a set of propagating rays, reflecting, diffracting, and scattering over environment elements to obtain the accurate characterization and simulation of electromagnetic propagation. The parameters used to generate the dataset are listed in Tab. I.

Table I: Datasets parameters.

Parameters	Value	Parameters	Value
Carrier frequency	60GHz	BS antenna	[1,64,1]
Bandwidth	0.05GHz	UE antenna	[1,1,1]
Sub-carriers	64	Size of template	[8,16]

B. Visualization of Classification Areas

Firstly, we visualize the classification results of the scene as shown in Fig. 4. It can be seen that our pre classification has excellent effectiveness, dividing an outdoor scene with a large area of 250×250 meters and many buildings into lots of small areas with irregular boundaries. The data within each region has relatively similar distribution characteristics, and there is good distinguishability between regions as well, greatly reducing the data distribution heterogeneity problem. By leveraging these features, it can be used for channel estimation, channel inference, beamforming and so on.

C. Effectiveness of Classification Methods

We compare the proposed algorithm with the following baseline algorithms for NLOS when the antenna array is ULA.

1) Res_CFR/Res_ADCAM/Res_CFRADCAM: Use CFR or ADCAM or CFR and ADCAM together as input for the feature extractor, and then put the obtained features into a linear classifier to obtain the predicted coordinates.

2) Res_multi_CFRADCAM/Res_multi_CFRperfectADCAM: Compared to Res_CFRADCAM, these two baselines put the obtained features into multiple linear classifiers to obtain the predicted coordinates. Res_multi_CFRperfectADCAM replaces ADCAM with AOA, AOD, distance, gain of the first arrival path as input.

Fig. 5 presents compelling evidence of the efficacy of our methodology. Firstly, as epochs increase, the MSE for Res_ADCAM rapidly decreases, but ultimately close to that
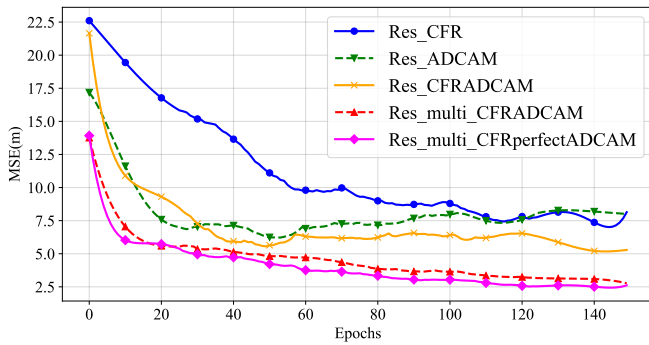
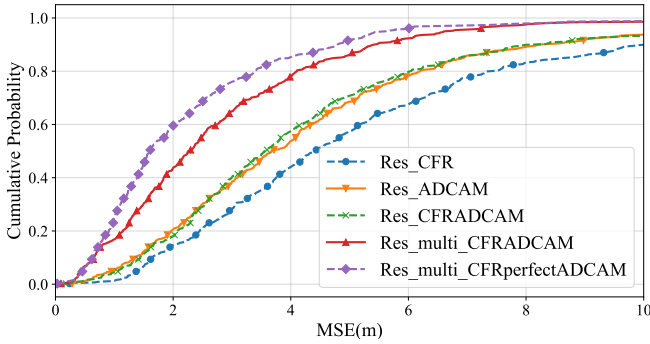Figure 5: MSE with epochs of multiple processing results.

Figure 6: CDF curves of multiple processing results.

of Res_CFR which is the highest. However, by adopting our AMDNLoc which fuses ADCAM and CFR deeply, the optimal MSE can be improved from 6.34 to 2.14 meters. This is because ADCAM is inherently sparser than CFR, and relying solely on either fingerprint can't fully represent the multi-path features. What's more, AMDNLoc is more focused on the changes between the locations and the fingerprints with the similar data distribution, so it can better reflect the true nature of the location-fingerprint relationship. Further supporting our findings, Fig. 6 illustrates that over 60% of positioning errors in our model fall within a 2-meter range. This is a significant increase compared to the baseline models, which only demonstrate a 20% probability of positioning errors within the same range. These results unequivocally demonstrate the superiority of our integrated approach in achieving more accurate and reliable positioning in MIMO-OFDM systems.

V. CONCLUSION

In the paper, we proposed a novel multi-sources information fusion learning framework named AMDNLoc that makes full use of the multi-path feature across the frequency, power, angle and delay domain as fingerprints to tackle the inherent heterogeneity issue of fingerprint distributions. We explore a two-stage matched filter for PFCFR corresponding to the distribution of CFR, and fuse it with ADCAM classification region after centroid-based clustering method. What's more, a segment-specific linear classifier mechanism sharing the same feature extractor is utilized to build regression relationship between fingerprint and locations after eliminating the negative effect of the region covariant. Numerical experiments have shown that the AMDNLoc achieves SOTA results compared

with traditional convolutional neural networks on the WAIR-D.

REFERENCES

[1] L. Xiao, A. Behboodi, and R. Mathar, "Learning the localization function: Machine learning approach to fingerprinting localization," [Online] available: *https://arxiv.org/abs/1803.08153*, Mar. 2018.

[2] G. P. Bittencourt, A. A. Urbano, and D. C. Cunha, "A proposal of an rf fingerprint-based outdoor localization technique using irregular grid maps," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Jun. 2018.

[3] Q. D. Vo and P. De, "A survey of fingerprint-based outdoor localization," *IEEE Commun. Surv. Tutor.*, vol. 18, no. 1, pp. 491–506, Jun. 2015.

[4] X. Sun, X. Gao, G. Y. Li, and W. Han, "Single-site localization based on a new type of fingerprint for massive MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6134–6145, Mar. 2018.

[5] B. Peng, G. Seco-Granados, E. Steinmetz, M. Fröhle, and H. Wymeersch, "Decentralized scheduling for cooperative localization with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4295–4305, Apr. 2019.

[6] X. Sun, C. Wu, X. Gao, and G. Y. Li, "Fingerprint-based localization for massive MIMO-OFDM system with deep convolutional neural networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10 846–10 857, Sep. 2019.

[7] D. Li, Y. Lei, X. Li, and H. Zhang, "Deep learning for fingerprint localization in indoor and outdoor environments," *Int. J. Geoinf.*, vol. 9, no. 4, p. 267, Feb. 2020.

[8] A. Del Corte-Valiente, J. M. Gómez-Pulido, O. Gutiérrez-Blanco, and J. L. Castillo-Sequera, "Localization approach based on ray-tracing simulations and fingerprinting techniques for indoor–outdoor scenarios," *Energies*, vol. 12, no. 15, p. 2943, Jul. 2019.

[9] J. Gante, G. Falcao, and L. Sousa, "Deep learning architectures for accurate millimeter wave positioning in 5g," *Neural Proc. Lett.*, vol. 51, no. 1, pp. 487–514, Aug. 2020.

[10] X. Gong, X. Fu, X. Liu, and X. Gao, "Transformer-based fingerprint positioning for multi-cell massive mimo-ofdm systems," in *Proc. Int. Conf. Info. Edu. Tech. (ICIET)*, Mar. 2023.

[11] P. Ferrand, A. Decurninge, and M. Guillaud, "DNN-based localization from channel estimates: Feature design and experimental results," in *Proc. IEEE Global Commun. Conf. (Globecom)*, Dec. 2020.

[12] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv. (CSUR)*, vol. 46, no. 2, pp. 1–32, Dec. 2013.

[13] U. Bhattacherjee, C. K. Anjinappa, L. Smith, E. Ozturk, and I. Guvenc, "Localization with deep neural networks using mmwave ray tracing simulations," in *Proc. IEEE SoutheastCon*, Mar. 2020.

[14] F. Zhu, B. Wang, Z. Yang, C. Huang, Z. Zhang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Robust Millimeter Beamforming via Self-Supervised Hybrid Deep Learning," in *Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2023.

[15] Y. Huangfu, J. Wang, S. Dai, R. Li, J. Wang, C. Huang, and Z. Zhang, "WAIR-D: Wireless AI research dataset," [Online] available: *https://arxiv.org/abs/2212.02159*, Dec. 2022.

[16] B. Wang, "AMDNloc," https://github.com/Horizontal666/AMDNloc, 2023.

[17] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Compu. Vis. Pat. Recog. (CVPR)*, Jun. 2016.