electronics

MDPI

Article

Augmenting Beam Alignment for mmWave Communication Systems via Channel Attention

Jihyung Kim [1] **and Junghyun Kim** [2,3,*]

[1] Spatial Wireless Transmission Research Section, Electronics and Telecommunications Research Institute, Daejeon 34129, Republic of Korea; savant21@etri.re.kr
[2] Department of Artificial Intelligence, Sejong University, Seoul 05006, Republic of Korea
[3] Deep Learning Architecture Research Center, Sejong University, Seoul 05006, Republic of Korea
* Correspondence: j.kim@sejong.ac.kr

Abstract: The beamforming technique has attracted considerable attention in wireless communication due to its various advantages such as interference reduction and improved wireless resource efficiency. However, the beam alignment between transmitting and receiving devices, which is fundamentally required for the beamforming, poses a significant challenge due to the continuous variability of the wireless channel. Recently, a deep learning-based technique has been proposed to predict narrow beam indices by measuring wide beams. However, there is room for improvement in the performance of the neural network architecture employed in this technique. Therefore, we suggest a novel deep learning model architecture that incorporates a channel attention module for beam training. The simulation results show a significant enhancement in performance with our scheme compared to both a state-of-the-art approach and other existing methods across all scenarios. Particularly, we confirm that even when reducing the number of wide beams used for measurement by approximately 50%, our proposed approach achieves a performance close to that of the state-of-the-art scheme.

Keywords: channel attention; deep learning; mmWave communications; beam prediction

Citation: Kim, J.; Kim, J. Augmenting Beam Alignment for mmWave Communication Systems via Channel Attention. *Electronics* **2023**, *12*, 4318. https://doi.org/10.3390/electronics12204318

Academic Editor: Martin Reisslein

Received: 16 September 2023
Revised: 13 October 2023
Accepted: 17 October 2023
Published: 18 October 2023

1. Introduction

Millimeter-wave (mmWave) wireless communication systems [1–3] enable high data rate communication due to their large bandwidth resources. However, high-frequency wireless signals are subject to rapid attenuation with distance and are highly sensitive to obstacles. To overcome these challenges, directional communication utilizing the multiple-input multiple-output (MIMO) technology [4] is applied to ensure high beamforming gain. Conventional beamforming techniques depend on narrow beams that necessitate precise channel data and are susceptible to beam misalignment due to factors such as movement or changing wireless channel environmental conditions. Furthermore, creating and assessing a multitude of narrow candidate beams across the entire angular spectrum results in substantial training expenses. One simple and direct method to address this cost is to confine the beam search to those with a fixed angular separation. Subsequently, by using the diminished received signals, we can forecast the optimal narrow beam. Nonetheless, this technique may prove less efficacious in scenarios characterized by a low signal-to-noise ratio (SNR), particularly when the actual angles of arrival and departure of the primary path do not precisely align with the main lobe of the candidate beam.

In recent years, deep learning (DL) has garnered increasing focus as a promising technology for 6G, particularly in the field of wireless communications [5]. DL has demonstrated substantial progress with regard to both performance and efficiency, with notable contributions in various areas of wireless communication, including channel state information (CSI) feedback [6], channel denoising [7], channel decoding [8,9], end-to-end transceivers [10], and beam prediction [11–15]. This paper specifically focuses on DL-assisted beam prediction. In a prior work [11], a method was introduced to forecast the

best narrow beam by leveraging wide beam measurements to cover the entire angular space. The authors performed a comparative study that involved evaluating both super-resolution beam prediction and super-resolution image restoration techniques. During this investigation, they created a dedicated convolutional neural network (CNN) optimized for beam prediction.

Recently, Ma et al. [15] used CNN for beam prediction from wide beam signals but faced noise sensitivity. To address this, they added LSTM to track user movements and refine beam direction. However, the combined CNN-LSTM model still struggled to provide adequate beamforming gain. For example, there are areas where the combined CNN-LSTM model performs worse than the CNN model despite increased complexity. Therefore, we focused on the need for a more effective neural network architecture.

Since the introduction of AlexNet [16], neural networks have conventionally been constructed to be deep, enabling them to capture intricate characteristics with nonlinearities. Alternatively, there has been a surge in research focused on enhancing model efficiency using methods other than increasing depth. This trend has been spurred by the introduction of backbone structures such as DenseNet [17], ResNet [18], EfficientNet [19], and ParNet [20].

Inspired by these effective architectural innovations, we have devised a novel neural network structure based on attention mechanisms [21]. This architecture, named by channel attention, proficiently extracts and accentuates features from complex high-dimensional signals. In our experiments, conducted within the same framework as [15], we observed that integrating the channel attention module into the combined CNN and LSTM structure yields a substantial increase in beamforming gain.

Through experiments, we have confirmed that our proposed model demonstrates superior performance from a normalized beamforming gain perspective when compared to a state-of-the-art model and other existing models. Furthermore, when we applied the optimal neighboring criterion (ONC) and maximum probability criterion (MPC) techniques proposed in [15] to our model to reduce the wide beam measurement overhead, we observed that despite a roughly 50% reduction in the number of wide beams for beam measurement, our model's performance closely approximates that of the state-of-the-art model that utilizes all wide beams.

2. System Model

We focus on a downlink wireless communication system employing mmWave MIMO technology where base station (BS) and user equipment (UE) are equipped with M_{Tx} and M_{Rx} antennas, respectively. Within the scope of our assumptions, we contemplate a channel configuration where antennas are positioned in a two-dimensional layout, utilizing uniform linear arrays (ULAs).

2.1. Channel Model

We utilize a narrowband frequency-flat channel model featuring a line-of-sight (LOS) path along with C non-line-of-sight (NLOS) clusters. We use $\boldsymbol{H}_{\mathrm{LOS}}$ and $\boldsymbol{H}_{\mathrm{NLOS}}$ to represent the LOS and NLOS parts of the channel matrix $\boldsymbol{H} \in \mathbb{C}^{M_{\mathrm{Rx}} \times M_{\mathrm{Tx}}}$, respectively. Each part is formulated as follows:

$$\boldsymbol{H}_{\mathrm{LOS}} = \sqrt{\frac{M_{\mathrm{Tx}} M_{\mathrm{Rx}}}{\rho_{\mathrm{LOS}}}} \alpha_{\mathrm{LOS}} \mathbf{a}_{\mathrm{Rx}}(\theta_{\mathrm{LOS}}) \mathbf{a}_{\mathrm{Tx}}^{H}(\phi_{\mathrm{LOS}}), \tag{1}$$

$$\boldsymbol{H}_{\mathrm{NLOS}} = \sum_{c=1}^{C} \sqrt{\frac{M_{\mathrm{Tx}} M_{\mathrm{Rx}}}{\rho_c}} \sum_{l=1}^{L_c} \frac{\alpha_{c,l}}{\sqrt{L_c}} \mathbf{a}_{\mathrm{Rx}}(\theta_c + \theta_{c,l}) \mathbf{a}_{\mathrm{Tx}}^{H}(\phi_c + \phi_{c,l}). \tag{2}$$

The LOS path is characterized by the pathloss ρ_{LOS}, angle-of-arrival (AoA) θ_{LOS}, and angle-of-departure (AoD) ϕ_{LOS}. Similarly, the c-th cluster, which consists of L_c paths, is characterized by the pathloss ρ_c, AoA θ_c and AoD ϕ_c. For the l-th path within the c-th

cluster, we use $\alpha_{c,l}$, $\theta_{c,l}$, and $\phi_{c,l}$ to represent the complex gain, AoA offset, and AoD offset, respectively. The antenna response vectors are represented by

$$\mathbf{a}_{\mathrm{Tx}}(\phi) = \sqrt{\frac{1}{M_{\mathrm{Tx}}}}\left[1\ e^{j2\pi d_{\mathrm{Tx}}\sin(\phi)/\lambda}\ \ldots\ e^{j2\pi(M_{\mathrm{Tx}}-1)d_{\mathrm{Tx}}\sin(\phi)/\lambda}\right]^T, \tag{3}$$

$$\mathbf{a}_{\mathrm{Rx}}(\theta) = \sqrt{\frac{1}{M_{\mathrm{Rx}}}}\left[1\ e^{j2\pi d_{\mathrm{Rx}}\sin(\theta)/\lambda}\ \ldots\ e^{j2\pi(M_{\mathrm{Rx}}-1)d_{\mathrm{Rx}}\sin(\theta)/\lambda}\right]^T, \tag{4}$$

where d_{Tx} and d_{Rx} represent the antenna spacings, and λ stands for the wavelength. To simplify, we assume $d_{\mathrm{Tx}} = d_{\mathrm{Rx}} = \lambda/2$. The transpose and conjugate transpose operations are represented as $(\cdot)^T$ and $(\cdot)^H$, respectively.

2.2. Beam Training Model

We explore the utilization of analog beamforming with phase shifters, where the transmitting beam from the BS is represented by $\boldsymbol{f} \in \mathbb{C}^{M_{\mathrm{Tx}}\times 1}$, and the UE's receiving beam is represented as $\boldsymbol{w} \in \mathbb{C}^{M_{\mathrm{Rx}}\times 1}$. For m-th, the candidate transmitting beam and n-th candidate receiving beam are expressed as

$$\boldsymbol{f}_m = \sqrt{\frac{1}{M_{\mathrm{Tx}}}}\left[1\ e^{j\pi\sin(\gamma_{\mathrm{Tx},m})}\ \ldots\ e^{j\pi(M_{\mathrm{Tx}}-1)\sin(\gamma_{\mathrm{Tx},m})}\right]^T, \tag{5}$$

$$\boldsymbol{w}_n = \sqrt{\frac{1}{M_{\mathrm{Rx}}}}\left[1\ e^{j\pi\sin(\gamma_{\mathrm{Rx},m})}\ \ldots\ e^{j\pi(M_{\mathrm{Rx}}-1)\sin(\gamma_{\mathrm{Rx},m})}\right]^T, \tag{6}$$

where $m \in \{1, 2, \ldots, N_{\mathrm{Tx}}\}$ and $n \in \{1, 2, \ldots, N_{\mathrm{Rx}}\}$. $\gamma_{\mathrm{Tx},m}$ and $\gamma_{\mathrm{Rx},n}$ represent the beam directions of the m-th transmitting beam and the n-th receiving beam, respectively. To encompass the entire angular range, we take into consideration the scenario where the directions of the beam pairs for transmitting and receiving have been uniformly selected from predetermined intervals $(-\Gamma_{\mathrm{Tx}}/2, \Gamma_{\mathrm{Tx}}/2)$ and $(-\Gamma_{\mathrm{Rx}}/2, \Gamma_{\mathrm{Rx}}/2)$, i.e.,

$$\gamma_{\mathrm{Tx},m} = -\frac{\Gamma_{\mathrm{Tx}}}{2} + \frac{2m-1}{2N_{\mathrm{Tx}}}\Gamma_{\mathrm{Tx}}, \tag{7}$$

$$\gamma_{\mathrm{Rx},n} = -\frac{\Gamma_{\mathrm{Rx}}}{2} + \frac{2n-1}{2N_{\mathrm{Rx}}}\Gamma_{\mathrm{Rx}}. \tag{8}$$

Considering the channel matrix $\boldsymbol{H}$ and an associated beam pair $\{\boldsymbol{f}, \boldsymbol{w}\}$, we can represent the received signal y in the following manner:

$$y = \sqrt{P}\boldsymbol{w}^H\boldsymbol{H}\boldsymbol{f}x + \boldsymbol{w}^H\boldsymbol{n}, \tag{9}$$

where P is the transmission power, x is the transmitting signal with $|x| = 1$, and $\boldsymbol{n} \in \mathbb{C}^{M_{\mathrm{Rx}}\times 1}$ denotes AWGN vector with zero mean and variance of σ^2, i.e., $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}_{M_{\mathrm{Rx}}}, \sigma^2\boldsymbol{I}_{M_{\mathrm{Rx}}})$. To simplify the analysis, we focus on selecting the transmitting beam $\boldsymbol{f}$ at the BS side, assuming a single-antenna UE and omitting the receiving beam $\boldsymbol{w}$. Due to the limited penetration ability and substantial power loss through reflection in mmmWave channels, the power of the LOS path is significantly higher than that of NLOS paths. As a result, the LOS path exerts a predominant influence on mmWave channels, and the received signal components from the NLOS path can be considered as a form of noise. It can be represented by the following

$$y = \sqrt{P}\boldsymbol{H}_{\mathrm{LOS}}\boldsymbol{f}x + \sqrt{P}\boldsymbol{H}_{\mathrm{NLOS}}\boldsymbol{f}x + n \tag{10}$$

$$= \sqrt{P}\boldsymbol{H}_{\mathrm{LOS}}\boldsymbol{f}x + n_{\mathrm{eq}}, \tag{11}$$

where the composite noise term $n_{\mathrm{eq}} = \sqrt{P}\boldsymbol{H}_{\mathrm{NLOS}}\boldsymbol{f}x + n$.

The goal of beam training is to determine a beam pair $\{\boldsymbol{f}_{m^*}, \boldsymbol{w}_{n^*}\}$ that yields the highest gain. This objective can be formulated as the following optimization problem:

$$\{m^*, n^*\} = \underset{\substack{m\in\{1,2,\ldots,N_{\mathrm{Tx}}\},\\ n\in\{1,2,\ldots,N_{\mathrm{Rx}}\}}}{\arg\max} \left|\boldsymbol{w}_n^H \boldsymbol{H} \boldsymbol{f}_m\right|^2. \tag{12}$$

One straightforward approach to address the aforementioned optimization is to employ a brute-force search. In this scheme, all possible candidate beams are systematically explored to identify the beam pair that yields the highest power in the measured signal. Nonetheless, implementing this scheme necessitates a large number of measurements, specifically $N_{\mathrm{Tx}}N_{\mathrm{Rx}}$, which results in a substantial training overhead. In order to resolve this issue, we can investigate the utilization of a two-level beam search technique, incorporating a multi-resolution codebook. The codebook is organized in a hierarchical manner, with wide beam codewords at the primary level and narrow beam codewords at the subsequent level. This configuration offers a promising method for minimizing training overhead.

In [15], a narrow beam training scheme utilizaing a wide beam codebook is introduced. The received signal corresponding to the m-th wide beam is denoted as $\boldsymbol{y}_{\mathrm{w},m}$, and all received signals from the wide beams are combined into a single received signal vector $\boldsymbol{y}_{\mathrm{w}} = \left[\boldsymbol{y}_{\mathrm{w},1}\ \boldsymbol{y}_{\mathrm{w},2}\ \cdots\ \boldsymbol{y}_{\mathrm{w},N_{\mathrm{Tx}}/s_{\mathrm{Tx}}}\right]^T$ where s_{Tx} defines the number of narrow beams within each wide beam. Due to the increased angular resolution provided by the narrow beam codebook, the objective of the scheme is to forecast the index m^* of the best narrow beam based on the received wide beam signal vector. As the number of available candidate narrow beams is constrained, this prediction task can be viewed as a multi-classification problem, where each class represents a specific narrow beam. The prediction model is mathematically formulated as a function $f_1(\cdot)$, as described below:

$$m^* = f_1(\boldsymbol{y}_{\mathrm{w}}), \quad m^* \in \{1, 2, \ldots, N_{\mathrm{Tx}}\}. \tag{13}$$

Implementing this prediction model using conventional estimation methods is challenging for two main reasons. Firstly, the correlation between $\boldsymbol{y}_{\mathrm{w}}$ and ϕ_{LOS} demonstrates a strongly nonlinear behavior. Secondly, obtaining the distribution of the composite noise term n_{eq} is a demanding task due to the variability of NLOS paths in different propagation environments. These two factors contribute to the complexity of the estimation process using conventional methods. As a result, DL is employed, leveraging its strong capability to learn complex nonlinear relationships, in order to facilitate the implementation of the prediction task.

3. Calibrated Beam Training with Channel Attention

The authors in [15] employed an architecture comprising CNN and LSTM for the calibrated beam training (CBT). The CNN module was responsible for expanding the 2-dimensional signal into a 256-dimensional signal, thereby facilitating characteristic extraction. Meanwhile, the LSTM was utilized to handle continuous input signals, making the model robust to noise and enabling it to capture signal variations over time. However, our simulations indicated that the double-layered CNN module was inefficient in characteristic abstraction, and depending solely on the LSTM had limitations in addressing this issue. Consequently, we suggest a novel structure that efficiently captures characteristics.

The architecture of our model is illustrated in Figure 1. Following the t-th training session, the received wide beam signals $\{\boldsymbol{y}_{\mathrm{w},1}, \boldsymbol{y}_{\mathrm{w},2}, \ldots, \boldsymbol{y}_{\mathrm{w},t}\}$ undergo initial processing through preprocessing, CNN, and channel attention modules to extract relevant features from the received signals. Following that, the LSTM module leverages the received signals from both the ongoing and previous beam training sessions to further fine-tune the narrow beam direction. Finally, the output module generates probabilities $\{\hat{p}_{1,t}, \hat{p}_{2,t}, \ldots, \hat{p}_{N_{\mathrm{Tx}},t}\}$,

with the candidate beam exhibiting the highest probability being selected as the optimal choice.

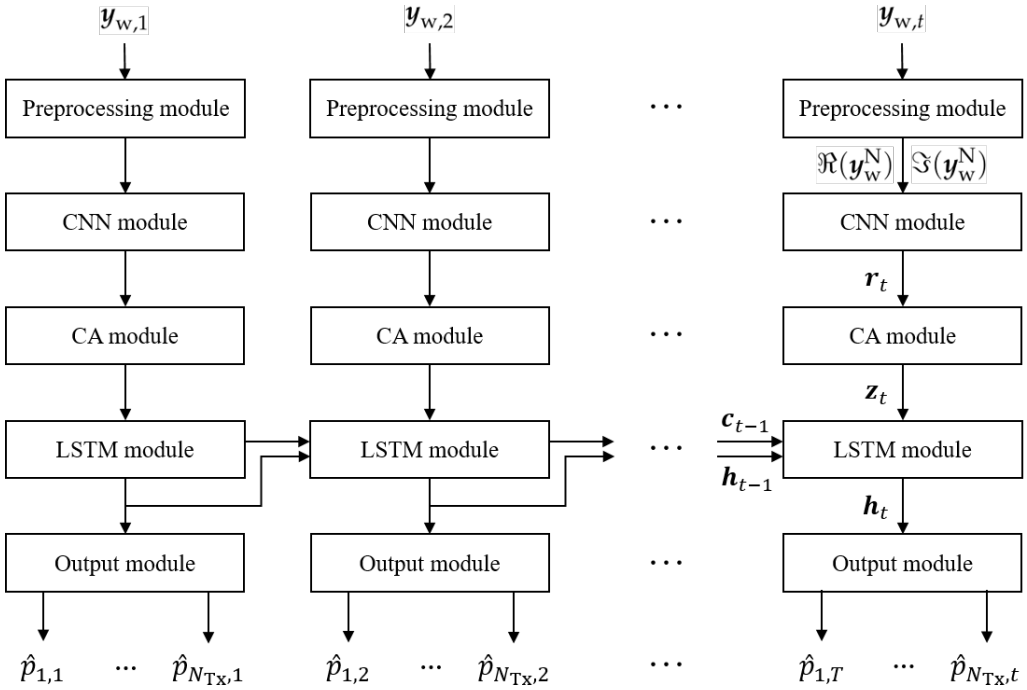

Figure 1. Proposed CNN-CA-LSTM model for the CBT scheme.

3.1. Preprocessing Module

Due to the complex-valued nature and large dynamic ranges of the received signal vector $\boldsymbol{y}_w$, it cannot be directly fed into the CNN module. To address this, the preprocessing module initially normalizes $\boldsymbol{y}_w$ by dividing it by the maximum magnitude of the elements. Mathematically, this is represented as follows

$$\boldsymbol{y}_w^N = \frac{\boldsymbol{y}_w}{\|\boldsymbol{y}_w\|_\infty}. \tag{14}$$

The normalized signal vector $\boldsymbol{y}_w^N = \Re(\boldsymbol{y}_w^N) + j\Im(\boldsymbol{y}_w^N)$ is split into the real term $\Re(\boldsymbol{y}_w^N)$ and the imaginary term $\Im(\boldsymbol{y}_w^N)$, which are provided as input to the subsequent convolution module.

3.2. Convolutional Neural Network Module

To derive hidden features from $\Re(\boldsymbol{y}_w^N)$ and $\Im(\boldsymbol{y}_w^N)$, multiple convolutional layers are employed. After each layer, a rectified linear unit (ReLU) activation layer is sequentially applied, enabling non-linear fitting capabilities. To prevent the model from becoming overly complex, a max-pooling is applied after the last ReLU activation. The max-pooling layer reduces each feature vector to a single scalar value through downsampling. The CNN module is depicted in Figure 2.

3.3. Channel Attention Module

A multi-scale feature map $\boldsymbol{r}$ obtained by the convolution module is considered as the input of our channel attention module. It goes through two different branches. On the left branch, shown in Figure 3, we use 1×1 convolution, followed by a batch normalization, to compress the channel-wise information, which can be represented as

$$\boldsymbol{u} = \mathrm{BN}(\mathrm{Conv}(\boldsymbol{r})). \tag{15}$$

On the right branch, a global average pooling (GAP) is employed to gather comprehensive spatial information on a global scale from the feature map $\boldsymbol{r}$. The k-th feature channel element of the pooled feature $\boldsymbol{g}$ is calculated as

$$g_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} r_k(i,j), \tag{16}$$

where $\boldsymbol{r}_k$ is the k-th feature channel element of the feature map $\boldsymbol{r}$. H and W denote the size of the vertical and horizontal dimensions of the feature map $\boldsymbol{r}$, respectively.

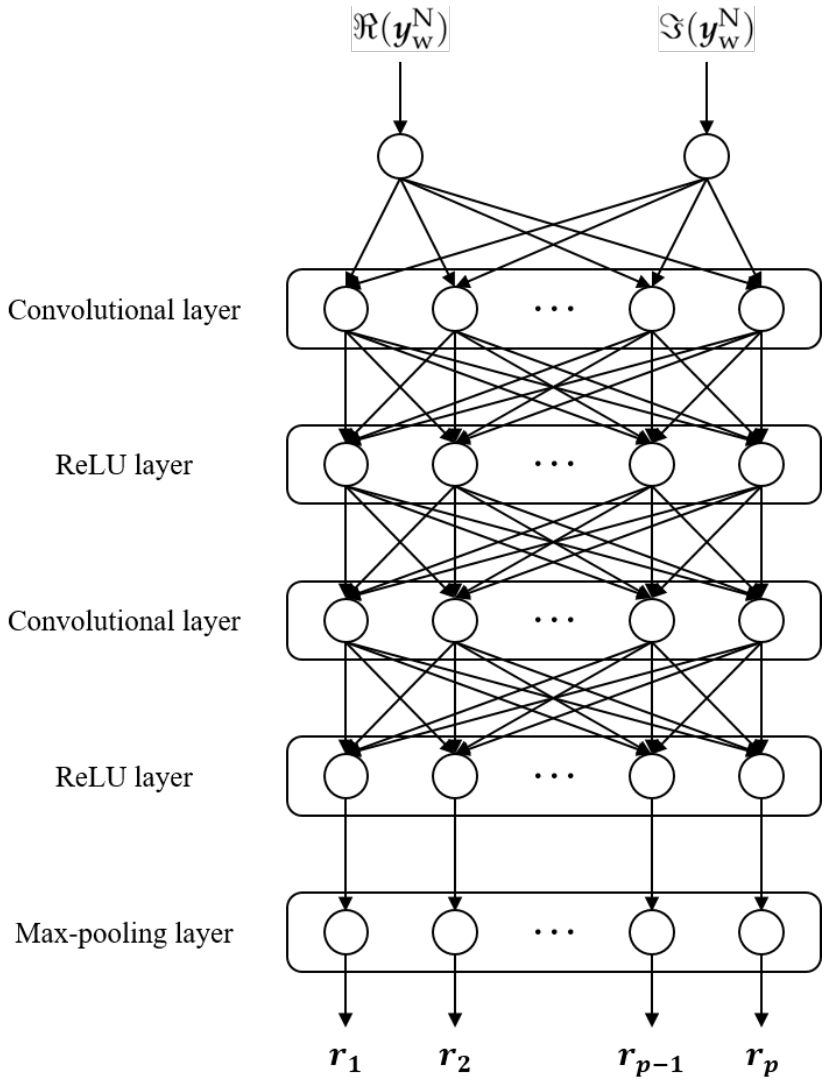

Figure 2. CNN module for the proposed CNN-CA-LSTM model.

The GAP compresses spatial dimensions and generates channel-wise statistics. We also connect 1×1 convolution with GAP for channel alignment with $\boldsymbol{u}$. An adaptive soft attention mechanism is employed across channels to flexibly determine different spatial scales, guided by the information in $\boldsymbol{g}$. To determine the importance of each spatial scale, a soft assignment weight is designed by

$$v = \mathrm{Sigmoid}(\mathrm{Conv}(\boldsymbol{g})). \tag{17}$$

Here, we apply the Sigmoid to re-calibrate the weight of the feature channel. Finally, we perform element-wise product $\odot$ on the refined weight $\boldsymbol{v}$ and the corresponding feature map $\boldsymbol{u}$, and apply the activation function Sigmoid-weighted linear unit (SiLU), as follows

$$\boldsymbol{z} = \mathrm{SiLU}(\boldsymbol{u} \odot \boldsymbol{v}), \tag{18}$$

where the SiLU, also known as the swish function, is composed of the original input multiplied by the Sigmoid function applied to the input, i.e., $\mathrm{SiLU}(x) = x \cdot \mathrm{sigmoid}(\beta x)$. In our model, we set $\beta = 1$.

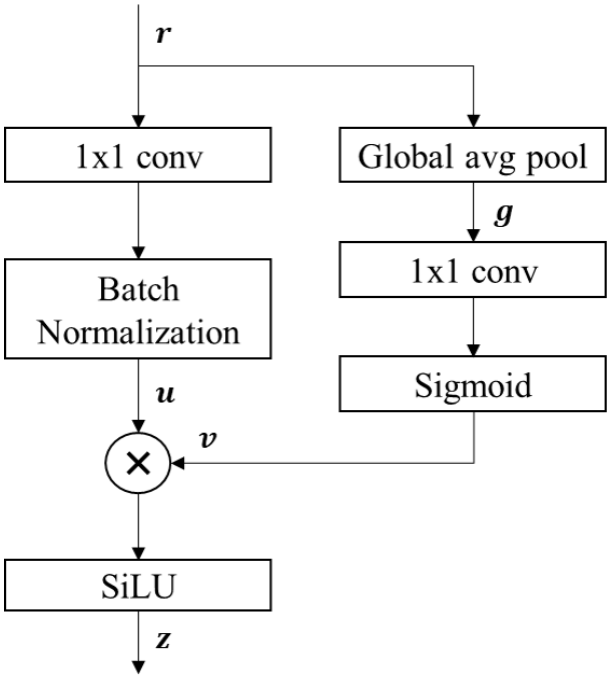

Figure 3. Proposed channel attention module.

3.4. Long Short-Term Memory Module

We present the formulation of the CBT scheme, which leverages the measured signals obtained from previous iterations of beam training. We make the assumption that the beam training process takes place at regular intervals, and we use the notation $\boldsymbol{y}_{\mathrm{w},t}$ to represent the measured signals from the t-th wide beam training.

To ascertain the best narrow beam m_t^* during the t-th wide beam training session, we leverage the received signals obtained from previous wide beam training sessions, denoted as $\{\boldsymbol{y}_{\mathrm{w},1}, \boldsymbol{y}_{\mathrm{w},2}, \ldots, \boldsymbol{y}_{\mathrm{w},t-1}\}$, and the current received signal $\boldsymbol{y}_{\mathrm{w},t}$. This task of prediction can be viewed as a problem of multi-class classification, employing a function $f_2(\cdot)$, as follows

$$m_t^* = f_2(\{\boldsymbol{y}_{\mathrm{w},1}, \boldsymbol{y}_{\mathrm{w},2}, \ldots, \boldsymbol{y}_{\mathrm{w},t}\}),\ m_t^* \in \{1, 2, \ldots, N_{\mathrm{Tx}}\}. \tag{19}$$

Since the AoD of the LOS path ϕ_{LOS} changes non-linearly with the movement of a UE, we employ a DL model to facilitate the prediction task. In contrast to the CNN-based model proposed in [15], in the LSTM-based model, the received signals from consecutive wide beam trainings and their corresponding best narrow beam indices for the UE are stacked in chronological order. This arrangement allows us to extract features related to the movement of the UE and forms a training sample.

Within the LSTM module, the current time slot input $\boldsymbol{z}_t$ is combined with the cell state and output $\{\boldsymbol{c}_{t-1}, \boldsymbol{h}_{t-1}\}$ from the previous time slot. This joint input is then passed through the LSTM at time slot t, allowing the LSTM to capture and learn features from previous inputs. The LSTM module generates characteristic features for optimal narrow

beam prediction by utilizing all successive received values and passes them to the output module. The LSTM structure for the proposed model is described by Figure 4. In the figure, $\sigma(\cdot)$ and $\tanh(\cdot)$ represent the sigmoid and hyperbolic activation functions, respectively. Among the two outputs of LSTM module, $\boldsymbol{c}_t$ and $\boldsymbol{h}_t$, only $\boldsymbol{h}_t$ serves as the input to the output module.

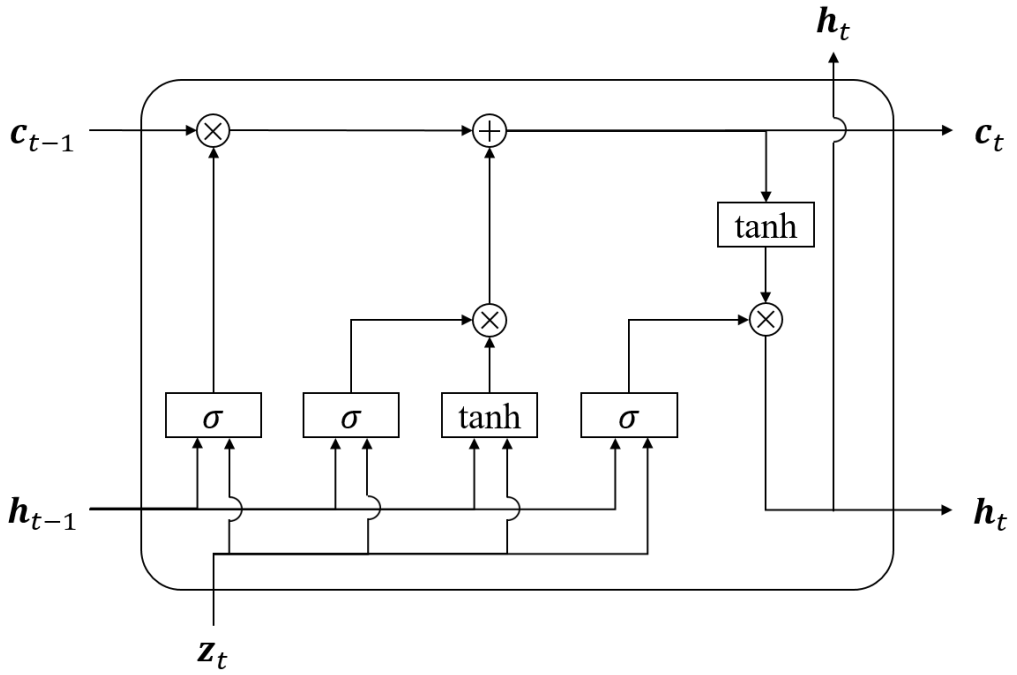

Figure 4. Basic LSTM structure for the proposed CNN-CA-LSTM model.

3.5. Output Module

To forecast the most suitable narrow beam among the candidate beams, we incorporate a fully connected (FC) layer after the LSTM module. This FC layer is responsible for converting the extracted features into representations that correspond to the candidate beams. Subsequently, to ensure that the outputs are in the form of probabilities, a softmax activation function is utilized to normalize the output values. Mathematically, this can be expressed as follows

$$\hat{\boldsymbol{p}} = \text{Softmax}(\text{FC}(\boldsymbol{h})), \tag{20}$$

where $\hat{\boldsymbol{p}}$ is a probability vector that signifies the probabilities associated with each beam being the optimal beam, while $\boldsymbol{z}$ corresponds to the output vector generated by the LSTM module. Finally, we select the narrow beam with the maximum probability as the ultimate choice, i.e.,

$$\hat{m}^* = \underset{m\in\{1,2,\ldots,N_{\text{Tx}}\}}{\arg\max}\ \hat{p}_m, \tag{21}$$

where $\hat{p}_m$ denotes the probability of the m-th candidate beam being the best choice.

We use a cross entropy function that is a commonly used cost function in classification tasks. It can be represented by

$$\text{Cost} = -\sum_{m=1}^{N_{\text{Tx}}} p_m \log \hat{p}_m, \tag{22}$$

where the value of p_m is equal to 1 when the m-th narrow beam is predicted as the best, and it is 0 otherwise.

4. Adaptive Calibrated Beam Training with Channel Attention

Following a methodology akin to the one proposed in [15], we introduce a channel attention-assisted adaptive CBT scheme aimed at mitigating the overhead associated with wide-beam measurements. In this scheme, we selectively train a subset of wide beams, utilizing received signals from previous beam training sessions. To initially ascertain the AoD of the LOS path, denoted as ϕ_{LOS}, across the entire angular space, we conduct an initial wide-beam training session where signals from all candidate wide beams are measured. Following this, only a subset of wide beams necessitates training, and we utilize the corresponding received signals to forecast the best narrow beam index.

In contrast to the adaptive CBT model introduced in [15], we have introduced a new channel attention module. The architecture of our channel attention-based adaptive CBT model is depicted in Figure 5. This structure effectively extracts the characteristics of wide beams from received signals, enabling the model to maintain satisfactory narrow beam prediction performance even when the number of wide beams used for measurement is reduced, as in partial beam training environments.

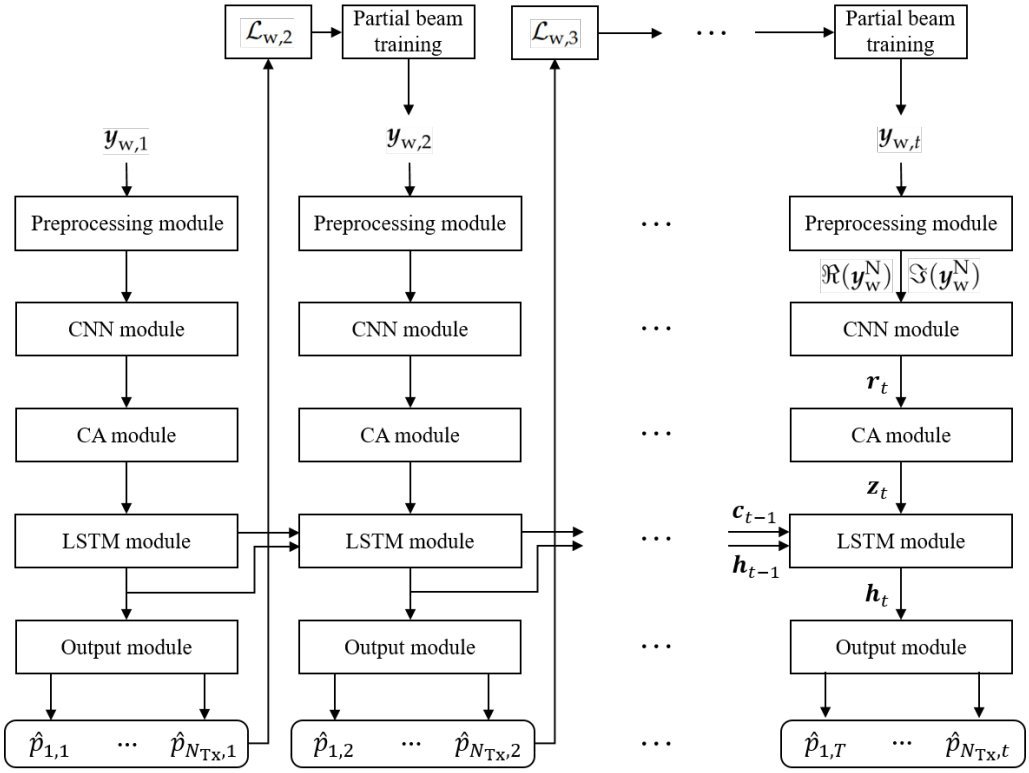

Figure 5. Proposed adaptive CNN-CA-LSTM model for the CBT scheme.

In order to identify candidate wide beams in the partial beam training, we employ two criteria introduced in [15], ONC and MPC. Let K represent the number of wide beams used for measurements during the t-th wide beam training session, where $K < N_{\mathrm{Tx}}/s_{\mathrm{Tx}}$ and $t > 1$. We denote the corresponding indices as $\mathcal{L}_{\mathrm{w},t}$.

The ONC's objective is to choose wide beams whose directions are closest to $\phi_{\mathrm{LOS},t}$. However, $\phi_{\mathrm{LOS},t}$ cannot be precisely determined. To approximate $\phi_{\mathrm{LOS},t}$ suitably, we take note that the UE's location during t-th wide beam training session is in close proximity to the location associated with the previous $(t-1)$-th wide beam training session. Consequently, we suggest approximating $\phi_{\mathrm{LOS},t}$ by utilizing the direction of the previously predicted optimal narrow beam.

The MPC relies on the assumption that the predicted probabilities reflect the quality of the beams. It chooses wide beams with the highest probabilities from the $(t-1)$-th prediction. It is important to note that the prediction results offer probabilities only for

narrow beams, not wide beams. To approximate the probability for the m-th wide beam, we aggregate the probabilities of all narrow beams that fall within the m-th wide beam.

5. Enhanced Adaptive Calibrated Beam Training with Channel Attention

The adaptive CBT scheme determines which wide beams to train based on the outcomes of prior beam predictions. However, this approach may become outdated in mobile scenarios if the AoD of the LOS path, denoted as ϕ_{LOS}, undergoes variations. To address this challenge and effectively track the changing AoD of the LOS path ϕ_{LOS}, we can leverage the received signals from previous beam training sessions to forecast the present UE location ahead of time. This allows us to calibrate the choice of wide beams to be trained, thereby further enhancing received SNRs. This scheme is called enhanced adaptive CBT [15].

In this section, we suggest an enhanced adaptive CBT scheme with channel attention, building upon our previously introduced channel attention approach. The structure of our channel attention-based enhanced adaptive CBT model is illustrated in Figure 6. Our attention structure effectively extracts the characteristics of wide beams, positively influencing the training of the auxiliary LSTM module for wide beam prediction. These advantages contribute to an enhancement in the final narrow beam prediction performance.

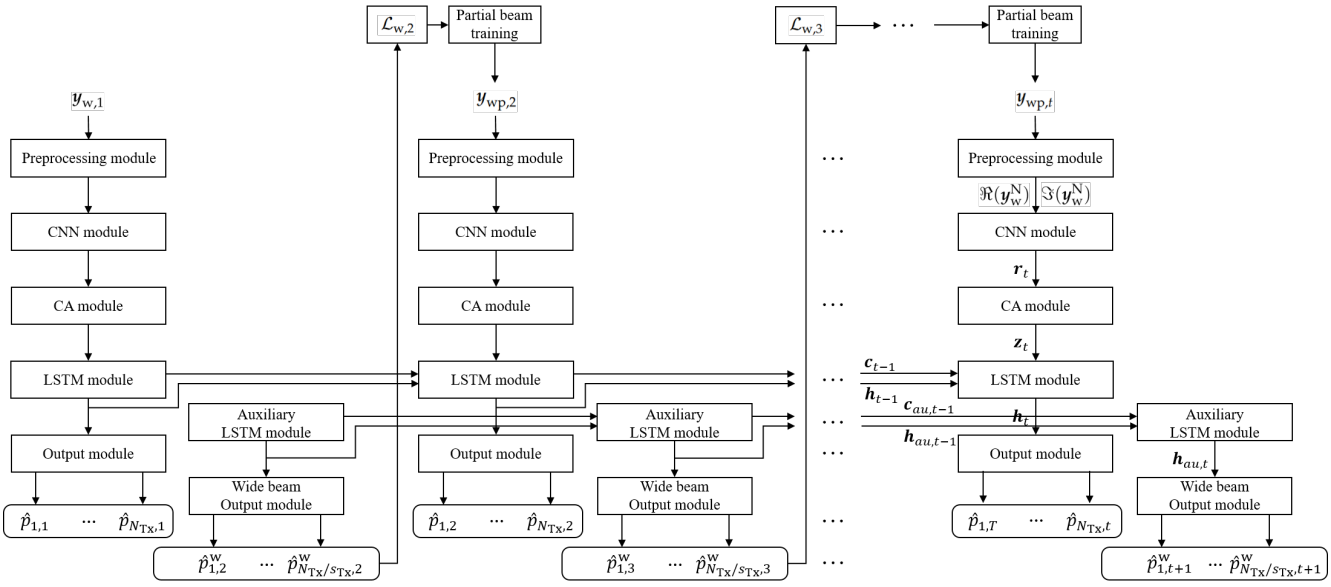

Figure 6. Proposed enhanced adaptive CNN-CA-LSTM model for the CBT scheme.

Specifically, in the channel attention-assisted enhanced adaptive CBT scheme, to identify wide beams for the t-th wide beam training session, an auxiliary LSTM module is utilized. The auxiliary LSTM module anticipates the t-th optimal wide beam index ahead of time, leveraging received signals from previous wide beam training sessions denoted as $\{\boldsymbol{y}_{\text{w},1}, \boldsymbol{y}_{\text{wp},2}, \ldots, \boldsymbol{y}_{\text{wp},t-1}\}$. The output of the auxiliary LSTM model is represented as predicted with wide beam probabilities, denoted as $\{\hat{p}^{\text{w}}_{1,t}, \hat{p}^{\text{w}}_{2,t}, \ldots, \hat{p}^{\text{w}}_{N_{\text{Tx}}/s_{\text{Tx}},t}\}$, with the wide beam having the highest predicted probability being determined as the best wide beam.

In order to train the entire model, including the auxiliary LSTM, we combine the costs of both narrow beam predictions, denoted as Cost_{n}, and wide beam predictions, denoted as Cost_{w}, using a weight coefficient μ. This combination is expressed as $\text{Cost} = \text{Cost}_{\text{n}} + \mu\text{Cost}_{\text{w}}$. The coefficient μ should be set to an appropriate value depending on how much the wide-beam prediction performance contributes to the final narrow-beam prediction. This can be optimized through sufficient preliminary experiments and performance analysis. In this paper, we set the value to 1. Within the enhanced adaptive CBT scheme, the two criteria, ONC and MPC, can also be applied, similar to the adaptive CBT scheme.

6. Numerical Results

We carried out simulations utilizing the COST2100 channel model [22], following the methodology and system parameters employed in [15]. We generated a dataset comprising 20,480 samples, with 80% dedicated to training, while the remaining 20% is reserved for testing. In our evaluation, we included the CNN-based model and the CNN-LSTM hybrid model introduced in [15], along with three baseline models: (1) A narrow beam prediction scheme [12] relies on calculating the ratio of the beamforming gains between the chosen wide beam and its adjacent wide beams; (2) A FC neural network-based beam prediction scheme [13], which utilizes $N_{\mathrm{Tx}}/s_{\mathrm{Tx}}$ measurements of sparsely sampled narrow beams; (3) Active learning-based beam prediction scheme [14], making use of $N_{\mathrm{Tx}}/s_{\mathrm{Tx}}$ measurements of hierarchical beams with a target resolution of N_{Tx}.

The detailed structure of our model is specified in Table 1. In the table, l_i denotes the count of input feature channels, while l_o signifies the count of output feature channels. In the convolution layers, the parameters $(\mathrm{p}_1, \mathrm{p}_2, \mathrm{p}_3)$ correspond to the filter size, stride size, and padding size, respectively. We denoted batch normalization as BN and dropout as DO at each layer. Throughout the training phase, the model undergoes a total of 80 epochs. The Adam optimizer is employed to optimize the trainable parameters.

Table 1. Detailed structure of the proposed models.

Module	Layer	Parameters
CNN	Convolution	$l_i = 2, l_o = 64, (3,3,1)$, BN
	ReLU	$l_i = 64, l_o = 64$
	Convolution	$l_i = 64, l_o = 256, (3,1,1)$, BN
	ReLU	$l_i = 256, l_o = 256$
	Max-pooling	$l_i = 256, l_o = 256$
Channel attention	Convolution	$l_i = 256, l_o = 256, (1,1,0)$, BN
	Global average pooling	$l_i = 256, l_o = 256$
	Convolution	$l_i = 256, l_o = 256, (1,1,0)$
	Sigmoid	$l_i = 256, l_o = 256$
	Multiplication	$l_i = (256,256), l_o = 256$
	SiLU	$l_i = 256, l_o = 256$
LSTM	LSTM	$l_i = 256, l_o = 256$, DO $= 0.2$
	LSTM	$l_i = 256, l_o = 256$, DO $= 0.2$
Auxiliary LSTM	LSTM	$l_i = 256, l_o = 256$, DO $= 0.2$
	LSTM	$l_i = 256, l_o = 256$, DO $= 0.2$
Narrow beam output	FC	$l_i = 256, l_o = 64$, DO $= 0.3$
	Softmax	$l_i = 64, l_o = 64$
Wide beam output	FC	$l_i = 256, l_o = 16$, DO $= 0.3$
	Softmax	$l_i = 16, l_o = 16$

Initially, we explored the influence of the number of training sessions t for wide beams on the achievable normalized beamforming gain G_{N}, as given in Figure 7. The normalized beamforming gain G_{N} is defined as

$$G_{\mathrm{N}} = \frac{|\boldsymbol{H}\boldsymbol{f}_{\hat{m}^*}|^2}{|\boldsymbol{H}\boldsymbol{f}_{m^*}|^2}, \tag{23}$$

where $\boldsymbol{f}_{\hat{m}^*}$ and $\boldsymbol{f}_{m^*}$ are the actual best narrow beam and the predicted best narrow beam, respectively. We consider 16 candidates for wide beams, which corresponds to $N_{\mathrm{Tx}}/s_{\mathrm{Tx}} = 16$. It is important to note that the CNN-assisted CBT scheme and all three baseline schemes remain independent of prior information, resulting in a constant beamforming gain regardless of t. The CNN-LSTM assisted CBT scheme demonstrates improved performance with increasing t. However, when $t = 1$, it demonstrates underwhelming performance compared to the CNN-assisted CBT. This suggests that there are limitations to enhancing

performance solely through the LSTM structure. On the contrary, our scheme consistently achieves the highest performance across all training numbers, thanks to the introduced channel attention module.

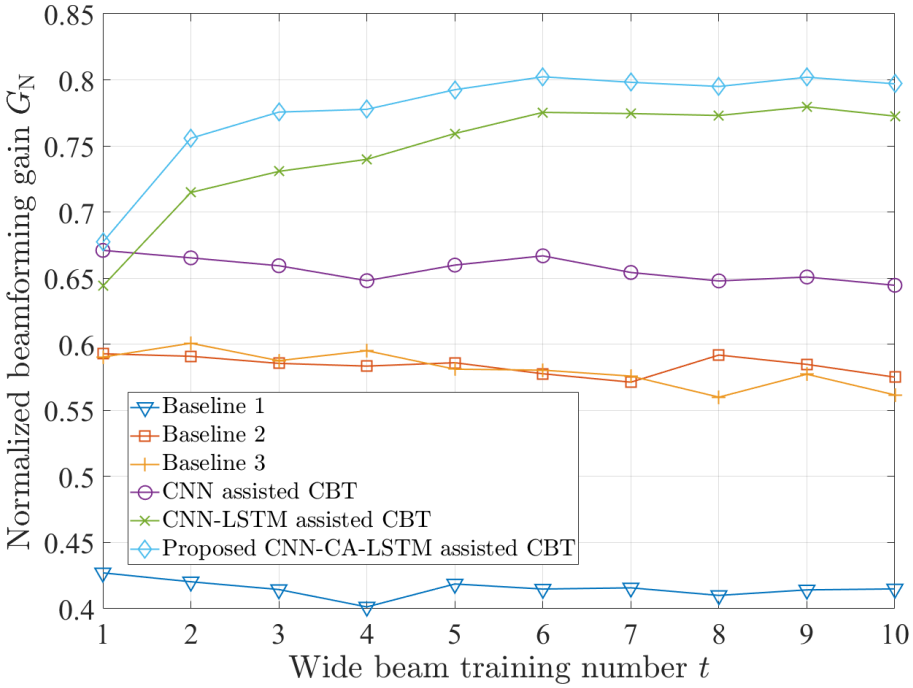

Figure 7. Normalized beamforming gain for various CBT schemes; baseline 1 [12], baseline 2 [13], baseline 3 [14], CNN assisted CBT [15], CNN-LSTM assisted CBT [15], and proposed CNN-CA-LSTM assisted CBT.

Figure 8 demonstrates that the effectiveness of our approach is confirmed by the cumulative distribution function (CDF) of the predicted narrow beam gain. As depicted in Figure 7, the normalized beamforming gains G_N of all methods converge when $t \leq 6$. Consequently, we rely on the G_N results after model convergence, specifically, utilizing the average G_N values over $t = 6$ to 10. As a result, Figure 8 also affirms the excellent prediction performance of our scheme in comparison to the five previous schemes.

Figure 9 illustrates the normalized beamforming gain G_N for adaptive CBT approaches. We set the candidate number of wide beams, K, used for partial beam training to 7. As anticipated, G_N exhibits an upward trend with increasing t for both ONC and MPC-based schemes, as a larger number of previously received signals offers more precise information regarding UE movement. Beyond $t = 6$, it becomes apparent that G_N stabilizes across all scenarios. After convergence, we verified that the performance of the proposed CNN-CA-LSTM assisted CBT scheme improved by approximately 3–4% compared to the performance of the conventional CNN-LSTM-assisted CBT scheme for both ONC and MPC methods. Furthermore, we confirmed that it achieves performance close to the conventional CNN-LSTM assisted CBT scheme, which uses 16 candidate beams, despite reducing the number of wide beams required for measurement for the narrow beam prediction from 16 to 7, resulting in an approximately 50% reduction.

Figure 10 shows the normalized beamforming gain G_N for enhanced adaptive approaches. In the experiment, we considered the case where the number of candidates K for the wide beam used for partial beam training is 7. Comparing the results in Figure 9, it can be observed that applying the enhanced adaptive approach to the proposed CNN-CA-LSTM-assisted CBT scheme yields better performance than applying the adaptive approach alone. This finding underscores the enhanced accuracy of UE location tracking in mobile scenarios achieved by the auxiliary LSTM. Notably, this performance enhancement becomes even more pronounced when the MPC technique is applied.

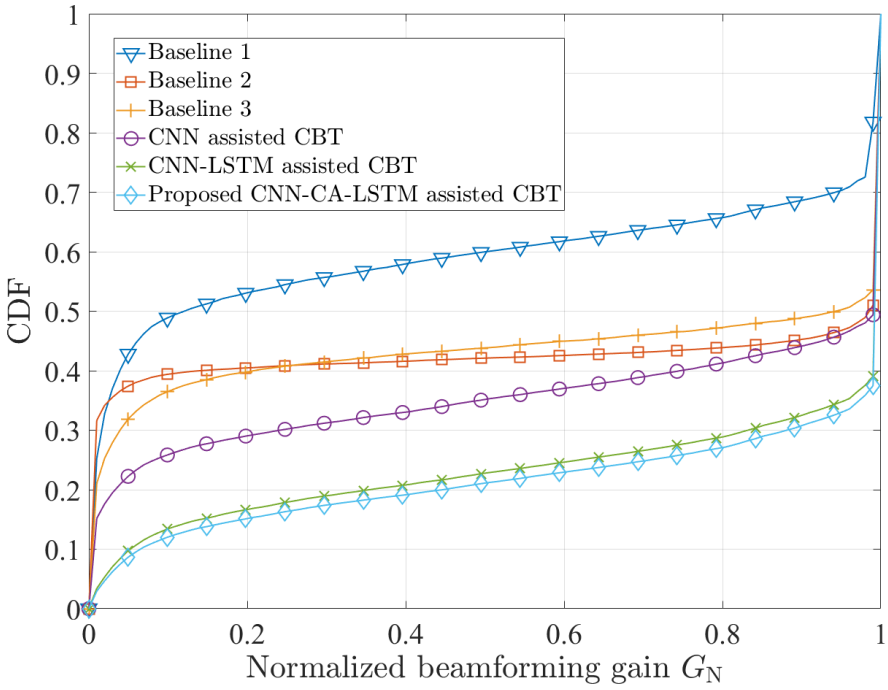

Figure 8. CDF of the predicted narrow beam gains for various CBT schemes; baseline 1 [12], baseline 2 [13], baseline 3 [14], CNN assisted CBT [15], CNN-LSTM assisted CBT [15], and proposed CNN-CA-LSTM assisted CBT.

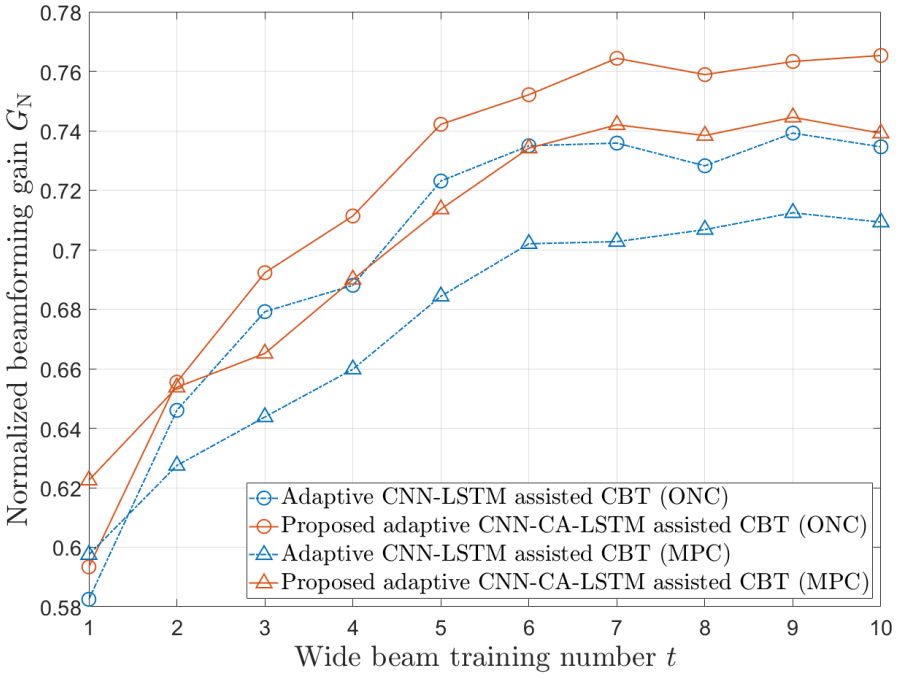

Figure 9. Normalized beamforming gain for adaptive CBT schemes; adaptive CNN-LSTM assisted CBT (ONC) [15], proposed adaptive CNN-CA-LSTM assisted CBT (ONC), adaptive CNN-LSTM assisted CBT (MPC) [15], and proposed adaptive CNN-CA-LSTM assisted CBT (MPC).

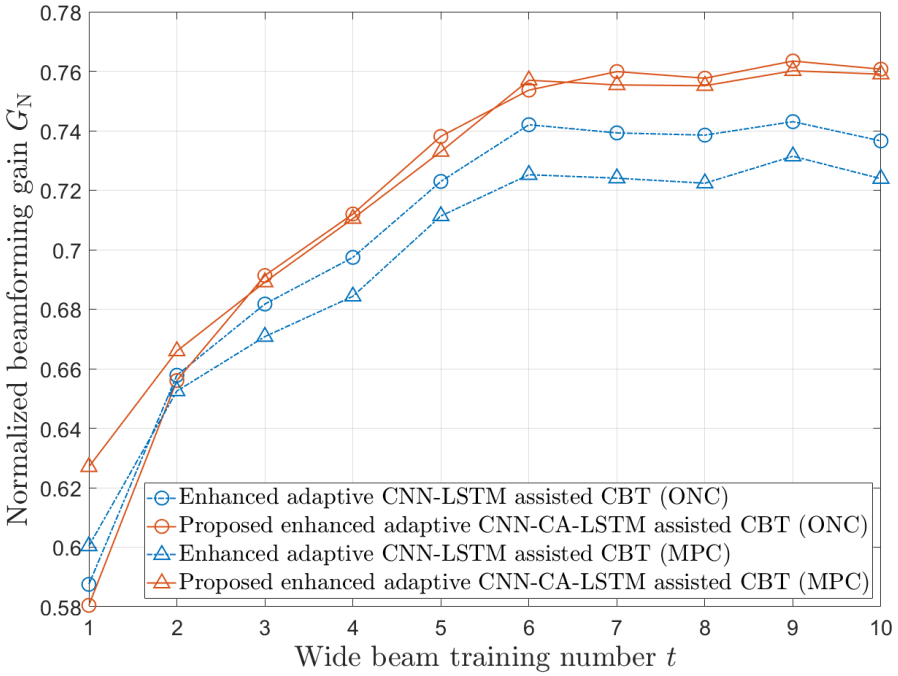

Figure 10. Normalized beamforming gain for enhanced adaptive CBT schemes; enhanced adaptive CNN-LSTM assisted CBT (ONC) [15], proposed enhanced adaptive CNN-CA-LSTM assisted CBT (ONC), enhanced adaptive CNN-LSTM assisted CBT (MPC) [15], and proposed enhanced adaptive CNN-CA-LSTM assisted CBT (MPC).

7. Conclusions

We conducted research on millimeter-wave beam alignment techniques using deep learning approaches for wireless communication systems. The incorporation of our proposed channel attention module has notably enhanced the performance of narrow beam forecasting by efficiently capturing characteristics from received wide beam signals. The simulations results consistently demonstrated that our CNN-CA-LSTM-assisted CBT scheme outperformed existing methods across various simulation environments. Moreover, the proposed CNN-CA-LSTM-assisted CBT scheme mitigates the performance degradation resulting from the application of two overhead reduction methods, ONC and MPC, to the state-of-the-art CNN-LSTM-assisted CBT scheme, ultimately achieving performance comparable that of the state-of-the-art scheme with full overhead.

Author Contributions: Conceptualization, J.K. (Jihyung Kim) and J.K. (Junghyun Kim); Methodology, J.K. (Jihyung Kim) and J.K. (Junghyun Kim); Software, J.K. (Jihyung Kim) and J.K. (Junghyun Kim); Validation, J.K. (Junghyun Kim); Writing—original draft, J.K. (Jihyung Kim); Writing—review and editing, J.K. (Junghyun Kim); Supervision, J.K. (Junghyun Kim). All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00436, Development of Standard Technologies for Reconfigurable Intelligent Surface Repeaters). This work was also supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the metaverse support program to nurture the best talents (IITP-2023-RS-2023-00254529) grant funded by the Korea government (MSIT).

Data Availability Statement: The data can be shared upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lin, Z.; Lin, M.; Champagne, B.; Zhu, W.-P.; Al-Dhahir, N. Secrecy-Energy Efficient Hybrid Beamforming for Satellite-Terrestrial Integrated Networks. *IEEE Trans. Commun.* **2021**, *69*, 6345–6360. [CrossRef]
2. An, K.; Lin, M.; Ouyang, J.; Zhu, W.-P. Secure Transmission in Cognitive Satellite Terrestrial Networks. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 3025–3037. [CrossRef]
3. Lin, Z.; An, K.; Niu, H.; Hu, Y.; Chatzinotas, S.; Zheng, G.; Wang, J. SLNR-Based Secure Energy Efficient Beamforming in Multibeam Satellite Systems. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 2085–2088. [CrossRef]
4. Lin, Z.; Niu, H.; An, K.; Wang, Y.; Zheng, G.; Chatzinotas, S.; Hu, Y. Refracting RIS-Aided Hybrid Satellite-Terrestrial Relay Networks: Joint Beamforming Design and Optimization. *IEEE J. Sel. Areas Inf. Theory* **2022**, *58*, 3717–3724. [CrossRef]
5. Lee, H.; Lee, B.; Yang, H.; Kim, J.; Kim, S.; Shin, W.; Shim, B.; Poor, H. V. Towards 6G Hyper-Connectivity: Vision, Challenges, and Key Enabling Technologies. *J. Commun. Netw.* **2023**, *25*, 344–354. [CrossRef]
6. Guo, J.; Wen, C.-K.; Jin, S.; Li, X. AI for CSI Feedback Enhancement in 5G-Advanced. *IEEE Wirel. Commun.* 2022, *early access*. [CrossRef]
7. Han, S.; Kim, J.; Song, H.-Y. A New Design of Channel Denoiser using Residual Autoencoder. *IET Electron. Lett.* **2023**, *59*, 1–3. [CrossRef]
8. Kim, H.; Oh, S.; Viswanath, P. Physical Layer Communication via Deep Learning. *IEEE J. Sel. Areas Inf. Theory* **2020**, *1*, 5–18. [CrossRef]
9. Choukroun, Y.; Wolf, L. Denoising Diffusion Error Correction Codes. *arXiv* **2022**, arXiv:2209.13533.
10. O'Shea, T.; Hoydis, J. An Introduction to Deep Learning for the Physical Layer. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 563–575. [CrossRef]
11. Echigo, H.; Cao, Y.; Bouazizi, M.; Ohtsuki, T. A Deep Learning-based Low Overhead Beam Selection in mmWave Communications. *IEEE Trans. Veh. Technol.* **2021**, *70*, 682–691. [CrossRef]
12. Luo, X.; Liu, W.; Wang, Z. Calibrated Beam Training for Millimeter-wave Massive MIMO Systems. In Proceedings of the IEEE Vehicular Technology Conference (VTC)-Fall, Honolulu, HI, USA, 22–25 September 2019.
13. Qi, C.; Wang, Y.; Li, G.Y. Deep Learning for Beam Training in Millimeter Wave Massive MIMO Systems. *IEEE Trans. Wirel. Commun.* 2020, *early access*. [CrossRef]
14. Chiu, S.-E.; Ronquillo, N.; Javidi, T. Active Learning and CSI Acquisition for mmWave Initial Alignment. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 2474–2489. [CrossRef]
15. Ma, K.; He, D.; Sun, H.; Wang, Z.; Chen, S. Deep Learning Assisted Calibrated Beam Training for Millimeter-wave Communication Systems. *IEEE Trans. Commun.* **2021**, *69*, 6706–6721. [CrossRef]
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012.
17. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
19. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019.
20. Goyal, A.; Bochkovskiy, A.; Deng, J.; Koltun, V. Non-deep Networks. *arXiv* **2021**, arXiv:2110.07641.
21. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
22. Liu, L.; Oestges, C.; Poutanen, J.; Haneda, K.; Vainikainen, P.; Quitin, F.; Tufvesson, F.; Doncker, P.D. The COST 2100 MIMO Channel Model. *IEEE Wirel. Commun.* **2012**, *19*, 92–99. [CrossRef]