Vision-assisted Beam Prediction for UAV-enabled Millimeter-Wave Communications using SE-ResNet50

Zahra Zarei
School of Electrical Engineering
Korea University
Seoul, Republic of Korea
zahraz@korea.ac.kr

Fitsum Debebe Tilahun
School of Electrical Engineering
Korea University
Seoul, Republic of Korea
fitsum_debebe@korea.ac.kr

Chung G. Kang
School of Electrical Engineering
Korea University
Seoul, Republic of Korea
ccgkang@korea.ac.kr

***Abstract*—The high overhead associated with the beam training process poses a significant challenge for highly mobile applications such as UAV communication. To mitigate this issue, this study proposes Squeeze-and-Excitation (SE) network that leverage visual information for accurate beam prediction in mmWave UAV communication. The SE network can selectively emphasize informative features through channel-wise feature recalibration, which enables the network to adapt to changing conditions, and optimize its predictions for different scenarios.**

***Index Terms*—UAV, mmWave, vision-assisted beam prediction, ResNet50, SE-ResNet50.**

I. Introduction

Unmanned aerial vehicles (UAVs), are widely recognized as a key enabler for the development of next-generation aerial networks and are expected to play a critical role in realizing advanced applications. To meet the demanding data rate requirements of these applications, equipping UAVs with mmWave transceivers and deploying large antenna arrays is indispensable. However, adjusting the narrow beams of these arrays, which is vital for ensuring a satisfactory signal-to-noise ratio, involves a significant training overhead that scales with the number of antennas. Moreover, the frequent updates required to maintain the optimal beam index due to the highly mobile nature of UAVs and their three-dimensional motion further exacerbate the beam training overhead. Therefore, it is imperative to explore novel approaches that can overcome these challenges and enable highly mobile mmWave UAV communication.

Various solutions have been proposed to address the optimal beam selection issue, focusing on beam training, channel estimation, and tracking. Recently, machine learning (ML) has gained attention for leveraging additional wireless environment information. For example, positional information is used to predict the optimal beam indices at the base station in [1], and [2], while visual data captured by cameras for beam prediction is employed in [3] and [4]. However, these solutions are based on a synthetic data and are geared towards scenarios where the users typically move in easy-to-predict mobility patterns in two dimensions. Recently, [5] proposed deep learning models for beam prediction utilizing sensory data from a large-scale real-world dataset, known as DeepSense 6G. In [5], deep learning models were introduced for beam prediction, using sensory data from a real-world dataset called DeepSense 6G [6]. In particular, the vision-assisted beam prediction task used the ResNet50 model. Inspired by this work, we propose a Squeeze-and-Excitation (SE) network to improve the accuracy of ResNet50 in beam prediction for vision-assisted tasks. The accuracy of the algorithm is evaluated based on the number of times the best beam pair is among the top-k predicted candidates, with k significantly smaller than the total number of beam pairs. To assess the performance of the proposed method, simulation data is generated using DeepSense 6G dataset [5].

This paper is organized as follows. Section II describes the system model. In Section III, we introduce the dataset, and the vision assisted beam prediction approaches. Simulation results and conclusion are presented in Section IV and V.

II. System Model and Problem Formulation

We consider a communication system in which a flying UAV with a single-antenna transmitter is served by a base station equipped with an M-element uniform linear array (ULA) and an RGB camera, Fig. 1. The communication system utilizes OFDM transmission with K subcarriers and a cyclic prefix of length D. To accommodate the mobile user, the base station is assumed to utilize a predefined beamforming codebook, $F = \{\mathbf{f}_q\}_{q=1}^{Q}, \mathbf{f}_q \in \mathbb{C}^{M\times 1}$, where Q represents the total number of beamforming vectors. Denoting the channel between the base station and the UAV at the k-th subcarrier at time t as $\mathbf{h}_k[t] \in \mathbb{C}^{M\times 1}$, the received signal at the UAV can be represented as

$$y_k = \mathbf{h}_k^T[t]\mathbf{f}_q[t]x + v_k[t] \tag{1}$$

where $v_k[t]$ is an additive Gaussian noise with mean of zero and variance of σ^2 [6]. The transmitted complex symbol x must comply with the constraint of $E[|x^2|] = P$, where P denotes the average symbol power. The beamforming vector

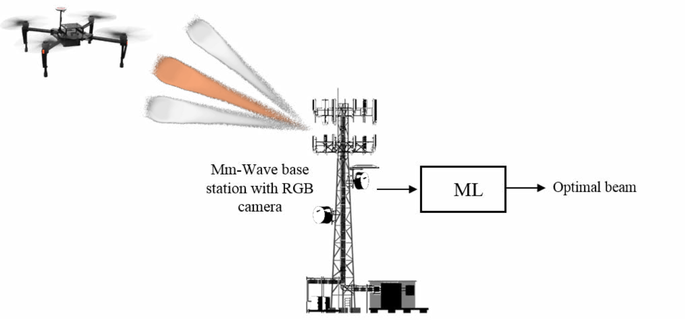

Fig. 1. System Model

$\mathbf{f}^*[t] \in \mathbf{F}$ at each time step t is chosen in a way to optimize the average received signal-to-noise ratio, $SNR = P/\sigma^2$, and can be expressed as

$$\mathbf{f}^*[t] = \arg\max_{\mathbf{f}_q \in \mathbf{F}} \frac{1}{K} \sum_{k=1}^{K} SNR |\mathbf{h}_k^T[t] \mathbf{f}_q[t]|^2 \tag{2}$$

Let $X\,[t] \in \mathbb{R}^{W \times H \times C}$ denote the RGB image captured by a camera installed in the base station at time step t, where W, H and C are the width, height, and the number of color channels of the image. Then, we intend to design a mapping function $f_\theta : X[t] \rightarrow \hat{\mathbf{f}}[t]$ which utilizes and predict the optimal beam index . Our objective is to design an optimal mapping which maximizes the accuracy of predictions for all U samples within the data set D can mathematically be stated as

$$f^*_{\theta^*} = \arg\max_{f_\theta} \prod_{u=1}^{U} P(\hat{\mathbf{f}}_{\mathbf{u}} = \mathbf{f}^*_{\mathbf{u}} | X_u)) \tag{3}$$

where $\mathbf{f}^*_{\mathbf{u}}$ is the optimal beam index given the RGB image of the u-th sample, X_u , from the data set [6].

III. Vision-Assisted Beam Prediction: Solution Approaches

In this work, we utilize the publicly available scenario 23 of the DeepSense 6G dataset. Next, we present a brief description of the data set, and proceed to discuss solution approaches of the problem in (3).

A. Dataset: DeepSense6G

DeepSense 6G is a multi-modal dataset, which includes vision, Radar, LiDAR, and GPS data, collected from real-world scenarios for sensing-assisted wireless communication applications. Specifically, Scenario 23 of the dataset is intended for investigating high-frequency wireless communication applications with UAVs. The testbed employs a standard-resolution RGB camera and an mmWave phased array with 16 elements operating at the 60GHz-band, while a codebook of 64 pre-defined beams are defined. Moreover, the mmWave phased array and RGB camera are positioned at a height of approximately 1.5 meters from the ground level, facing towards the sky to increase the basestation's field-of-view (FoV). Furthermore, to increase the dataset's diversity, the UAV's flight is conducted at various heights, distances, and speeds relative to the basestation [5].

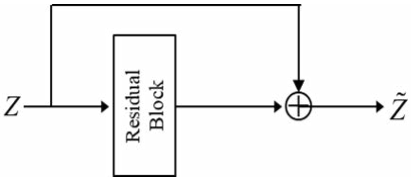

Fig. 2. Residual block with skip connection

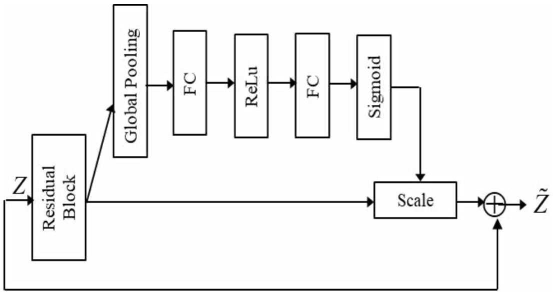

Fig. 3. Squeeze-and-excitation (SE) block

B. Baseline Approach

The baseline model, [6], adopts the vision-assisted beam prediction model presented in [5], which employs ResNet50, a popular convolutional neural network (CNN) architecture widely used for various computer vision tasks. It consists of 50 convolutional layers, including several residual blocks of different sizes. The residual blocks allow information to be directly passed from one layer to another without being transformed through the use of skip connections, as shown in Fig. 2. The use of these blocks has proven to be effective in mitigating the issue of vanishing gradients in very deep networks. The vision-assisted beam prediction model is trained and validated on the dataset. Moreover, the top-k accuracy is used as the metric to evaluate the proposed solution, which is defined as the percentage of test samples where the optimal ground-truth beam is included in the top-k predicted beams [6].

C. Proposed Solution Approach

The proposed vision-assisted beam prediction model integrates a Squeeze-and-Excitation (SE) network into ResNet50, to perform the classification task of mapping an image to a beam index. The Squeeze-and-Excitation (SE) network [7] is a novel deep learning architecture that seeks to enhance the efficacy of conventional CNNs models, such as ResNet50, by selectively highlighting informative features and attenuating less useful ones. The SE network achieves this goal by introducing a new module, commonly referred to as the SE block, into the traditional CNN architecture, as shown in Fig. 3. The SE block comprises two critical operations, namely, squeezing and exciting. During the squeezing operation, the SE block compresses the feature maps generated by the convolutional layers and Global Average Pooling (GAP) into a lower dimensional space that leads to generating channel wise statistics. Subsequently, during the exciting operation, the SE block applies a set of learned weights to the compressed feature maps, thereby enabling the network to accentuate or subdue significant features in a selective manner. To do so, the first FC layer with ReLU activation function $reduces$ the number of channel, while the second FC layer $expands$ the number of channels back to the original size of the input tensor. These FC layers serve as an attention mechanism that amplifies the important channels. Then, the Scale layer refers to a learnable parameter that scales the input features by passing the output of excitation network through a sigmoid activation function to obtain a scalar weight for each channel. These weights are multiplied with original input tensor to produce the output of SE block. This results in a refined set of feature maps that are more discriminative for the task [7].

The proposed squeeze-and-excitation ResNet50 (SE-ResNet50) capitalizes on the advantages of high and low-level features in the context of UAV movement, which requires both types of features due to its dynamic nature. In other words, the model can extract and integrate features at various levels of abstraction, thereby improving the accuracy and robustness of the model.

IV. Simulation Results

In this work, we utilize the publicly available scenario 23 of the DeepSense 6G dataset. The adopted testbed comprises of two units. Unit 1 primarily consists of a stationary base station equipped with an RGB camera and a mmWave phased array. The stationary unit adopts a 16-element 60GHz-band phased array and it receives the transmitted signal using an over-sampled codebook of 64 pre-defined beams. The camera is used to capture RGB images of 960 × 540 resolution at a base frame rate of 30 frames per second (fps). The RGB images are fed to the SE-Net consisting ResNet50 and fully connected (FC) layers.

As reported in Table I, in all scenarios the proposed network has outperformed the baseline approach in terms of top-k accuracy, achieving over 90% and near 100% accuracy for top-1, and top-5 beam selection, respectively. Moreover, since the speed of UAV and the height are two important factors in UAV communications, we simulated the results by grouping the speed and height into three different categories. In comparison with the baseline, the results show a noticeable improvement in terms of accuracy, over 81% and 82% accuracies in high-speed and low height flying UAV, respectively, which is a crucial in many problems. The reason for such improvement is that compared to ResNet50, SE blocks in SE-ResNet50 provide an additional layer of adaptivity that recalibrates feature responses dynamically based on the input, leading to selective amplification or suppression of specific features in a channel-wise manner.

TABLE I
Beam prediction results in terms of accuracy

	Scenarios	***Baseline Model***	***Proposed Model***
Top-k Accuracy	Top-1	86.32	90.831
	Top-2	97.12	98.912
	Top-3	99.41	99.699
	Top-5	99.69	99.935
Speed (S)	$S \leq 10$	86.2	90.800
	10 <S< 20	78.8	84.501
	$S \geq 20$	78.8	84.501
Height (H)	$H \leq 40$	78.1	82.105
	40 <H< 80	83.00	89.001
	$H \geq 80$	86.7	90.879

V. Conclusion

In conclusion, this paper presents a promising approach for accurate beam prediction in millimeter-wave (mm-Wave) communication systems (UAV) networks. We explored the utilization of image data captured by UAV-mounted cameras to overcome the challenges introduced by rapid channel variations due to UAV mobility.

This work proposed SE-ResNet50 for vision-assisted beam prediction for mmWave UAV communication. Simulation results demonstrate beam predication with SE-ResNet50 yields better accuracy for Top-1 beam selection in different scenarios as compared to the baseline ResNet50 approach. By adaptively adjusting the channel-wise contributions, excitation operation in the SENet effectively extracts features at different levels of abstraction. This fact enhances the network's ability to capture fine-grained details as well as high-level information, resulting in improved performance. In the future, it would be interesting to extend the current frameworks to make use of the remaining available sensory data in the data set.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under Grant 2020R1A2C100998413.

References

[1] Y. Wang, M. Narasimha, and R. W. Heath, "Mmwave beam prediction with situational awareness: A machine learning approach," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[2] S. Rezaie, C. N. Manchon, and E. de Carvalho, "Location- and orientation-aided millimeter wave beam selection using deep learning," in *IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[3] G. Charan, T. Osman, A. Hredzak, N. Thawdar, and A. Alkhateeb, "Vision-position multi-modal beam prediction using real millimeter wave datasets," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 2727–2731.

[4] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided 6G wireless communications: Blockage prediction and proactive handoff," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10193–10208, 2021.

[5] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided 6G wireless communications: Blockage prediction and proactive handoff," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10193–10208, 2021.

[6] G. Charan et al., "Towards Real-World 6G UAV Communication: Position and Camera Aided Beam Prediction," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Rio de Janeiro, Brazil, 2022, pp. 2951–2956.

[7] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 7132–7141.