Journal of Artificial Intelligence Practice (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2025.080117
ISSN 2371-8412 Vol. 8 Num. 1

An Improved Design of Beam Training Scheme in Vehicle Communication System

Wang Ruhang

Hefei University of Technology, Hefei, Anhui, China

Keywords: Artificial intelligence; Beam training; Vehicle communication system; Image enhancement

Abstract: This paper aims to study various technologies based on LIDAR, including distributed architecture, vision assisted wireless communication framework, multimodal machine learning framework, and self-calibrating illumination learning framework. The optimal algorithm will be selected and improved according to the specific situation of the actual dataset. The general plan is to first enhance low brightness images affected by weather and other environmental factors, and then use deep learning models to find the mapping relationship between visual image information and optimal codewords. In formal beam training, the accuracy of codeword prediction is improved to maximize the received power and improve the performance of vehicle networking communication.

1. Introduction

Millimeter wave communication technology plays a key role in improving driving safety and enhancing autonomous driving performance in vehicle networking communication systems. However, issues such as high-frequency signal attenuation and multipath effects require beam training, which involves directional transmission of beams to concentrate signal energy and maximize signal power. Therefore, it is particularly important to choose the best beam training scheme. The widely studied approach currently is codebook based beam training, which involves pre designing a set of beamforming vectors, where each vector corresponds to a specific direction of beam. The transmitting and receiving ends sequentially use the beams in the codebook for signal transmission. The receiving end measures the signal quality of each beam pair and provides feedback on the optimal index. Finally, the optimal combination of transmitting and receiving beams is selected based on the feedback. Compared to traditional beam training methods such as exhaustive scanning and layered scanning, codebook based beam training has the advantages of simple hardware implementation, strong compatibility, and widespread use in standardized protocols. However, the size of the codebook can affect training time, and in the case of a large number of antennas, it can also generate high time-frequency overhead. How to optimize the codebook structure and reduce the time-frequency overhead remains to be studied.

2. The trend of using artificial intelligence to improve telecommunications

2.1 The significance of beam training in mobile telecommunications

Beam training has emerged as a critical technology in modern communication systems, particularly in the context of millimeter-wave (mmWave) and terahertz (THz) frequencies. Millimeter wave and terahertz communications systems rely on the beamforming gains of the narrow beams to achieve sufficient receive signal power [1]. As the demand for higher data rates and more reliable connectivity grows, beam training plays a pivotal role in enabling efficient and robust wireless communication. One of the primary reasons beam training is essential is its ability to address the challenges of high-frequency signal propagation. LIDAR (light detection and ranging) is one of the most sophisticated sensors used in automated driving [2]. LIDAR data can be exploited without additional cost for improved communications when it is already used on an automated vehicle for mapping, positioning, or obstacle detection [3]. At mm Wave and THz frequencies, signals experience significant attenuation and are prone to obstacles, limiting their range and reliability. Beam training solves this problem by using directional antennas to focus signal energy in specific directions, thereby enhancing signal strength and extending coverage. As users move or obstacles alter signal paths, beam training continuously optimizes beam directions to maintain strong and stable connections. The increasing bandwidth and number of antennas do not come cheap. They bring with them a lot of control overhead that prevents them from realizing their full potential [4]. This adaptability ensures consistent performance, even in challenging scenarios such as mobile communication or indoor environments. Beam training also supports advanced technologies like massive Multiple-Input Multiple-Output (MIMO) and beamforming, which are integral to modern wireless systems. By precisely aligning beams between transmitters and receivers, it enhances the capacity and reliability of these systems, enabling them to support more users and deliver higher data rates. In conclusion, beam training is a cornerstone of modern communication, addressing the limitations of high-frequency signals, improving spectral efficiency, and enabling adaptive, high-performance networks.

2.2 The particular advance and convenience of artificial intelligence

Artificial Intelligence has become a transformative force in modern communication, revolutionizing how we connect, share information, and interact with technology. Its integration into communication systems has brought unprecedented efficiency, personalization, and innovation, making it a cornerstone of the digital age. One of the most significant contributions of AI in communication is its ability to optimize network performance. Through machine learning algorithms, AI can analyze vast amounts of data in real-time, predicting network congestion, identifying potential failures, and dynamically allocating resources. This ensures seamless connectivity, even in high-demand scenarios, such as crowded events or urban areas. For instance, AI-powered beamforming in 5G networks enhances signal strength and coverage, enabling faster and more reliable communication.AI also plays a pivotal role in personalizing user experiences. Virtual assistants like Siri, Alexa, and Google Assistant leverage natural language processing (NLP) to understand and respond to user queries, making interactions more intuitive and efficient. As AI continues to evolve, its impact on communication will only grow, shaping the future of how we connect and interact in an increasingly digital world. Thus, artificial intelligence plays an importance role in beam training. Developing solutions for the mmWave beam training and channel estimation overhead has attracted considerable interest over the last decade [5].

2.3 The current research status of image enhancement technology

In recent years, deep learning has revolutionized low-light image enhancement. Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) have become the cornerstone of modern techniques. These models are trained on large datasets of low-light and normal-light image pairs, enabling them to learn complex mappings between dark and bright images. For example, the EnlightenGAN and Zero-DCE (Zero-Reference Deep Curve Estimation) frameworks have demonstrated remarkable performance in enhancing low-light images while preserving natural colors and textures. Another promising direction is the integration of physical models with deep learning. By incorporating principles of image formation and light propagation, researchers have developed hybrid methods that combine the interpretability of traditional algorithms with the power of neural networks. These approaches often yield more robust and realistic results, particularly in challenging scenarios with extreme noise or uneven lighting. Despite these advancements, challenges remain. One major issue is the lack of high-quality datasets for training and evaluation. Many existing datasets are synthetic or lack diversity, limiting the generalization ability of models. Additionally, real-time performance is a critical requirement for applications like autonomous driving, where computational efficiency is paramount. Researchers are actively working on lightweight models and optimization techniques to address these concerns.

2.4 The motivation of developing the design of beam training scheme

The rapid development of vehicular communication systems, particularly in the context of the Internet of Vehicles (IoV), has created a pressing need for advanced technologies to ensure reliable, high-speed, and low-latency communication. Beam training has emerged as a key research area in this domain, driven by the unique challenges and requirements of modern vehicular networks. One of the primary motivations for studying beam training in vehicular communication systems is the increasing demand for high data rates and ultra-low latency. As vehicles become more connected and autonomous, they rely on real-time data exchange for applications such as collision avoidance, traffic management, and infotainment. Millimeter-wave and terahertz frequencies, which offer vast bandwidths, are essential to meet these demands. However, these high-frequency signals are highly susceptible to attenuation and blockages, making beam training crucial for maintaining strong and stable connections. Another critical factor driving research in beam training is the dynamic nature of vehicular environments. Vehicles move at high speeds, and their communication links are constantly affected by obstacles, reflections, and scattering. Traditional omnidirectional antennas are inefficient in such scenarios, as they waste energy and suffer from interference. Beam training enables directional communication by aligning narrow beams between vehicles and infrastructure, ensuring efficient signal transmission and reducing interference. Furthermore, the integration of massive Multiple-Input Multiple-Output and beamforming technologies in vehicular communication systems has highlighted the importance of beam training. These technologies rely on precise beam alignment to achieve high capacity and reliability. However, the fast-changing positions of vehicles and the need for real-time adaptation pose significant challenges. Research in beam training aims to develop algorithms and protocols that can quickly and accurately adjust beam directions, ensuring seamless communication even in highly mobile environments. Security and energy efficiency are additional motivations for advancing beam training in vehicular systems. By focusing signal energy only in the required direction, beam training reduces the risk of eavesdropping and minimizes power consumption, which is particularly important for battery-operated devices and energy-efficient networks.

3. The general idea of the design of beam training scheme

3.1 The design of low brightness image enhancement module

Firstly, we need to gather a dataset containing pairs of low-light and corresponding normal-light images. Public datasets like LOL (Low-Light) or custom datasets can be used. Then we need to apply transformations such as rotation, flipping, and noise addition to increase dataset diversity and improve model robustness. We also need to normalize pixel values to a range of [0, 1] or [-1, 1] to facilitate model training and split the dataset into training, validation, and test sets. Choosing a deep learning model suitable for image enhancement tasks is important. Popular choices include Convolutional Neural Networks, Generative Adversarial Networks , or specialized architectures like EnlightenGAN or Zero-DCE. It is vital to define an appropriate loss function to guide the training process. Common choices include Mean Squared Error for pixel-wise accuracy, perceptual loss for visual quality, and adversarial loss for GAN-based models. We need to use a deep learning framework like TensorFlow or PyTorch to implement the model. These frameworks provide pre-built layers and optimization tools that simplify development. We construct the model architecture by defining layers such as convolutional layers, activation functions, and upsampling layers. For GANs, we define both a generator and a discriminator network. We choose an optimizer like Adam or SGD to update model weights during training. We set appropriate learning rates and other hyperparameters. We implement a training loop that iterates over the dataset, computes the loss, and updates the model weights. We use mini-batch training to manage memory usage and improve convergence. We periodically evaluate the model on the validation set to monitor performance and avoid overfitting. We adjust hyperparameters if necessary. We save model checkpoints at regular intervals to allow for recovery in case of interruptions and to facilitate later evaluation. We use metrics like Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and visual inspection to assess the quality of enhanced images. We evaluate the trained model on the test set to measure its generalization ability. We compare results with baseline methods to demonstrate improvements. We implement an inference script that loads the trained model and processes new low-light images to produce enhanced versions. Optionally, we create a user-friendly interface using libraries like Tkinter or Flask to allow users to upload images and view enhanced results.

3.2 The design of multimodal fusion codeword prediction model

Designing a multimodal fusion codeword prediction model for multimodal fusion codeword prediction involves integrating diverse data sources, such as text, images, and audio, to predict codewords or labels that represent complex information. This task is particularly relevant in applications like multimedia analysis, autonomous systems, and human-computer interaction, where combining multiple modalities can significantly enhance prediction accuracy and robustness. At the core of the program is a deep learning model capable of processing and fusing multimodal data. The architecture typically begins with separate input branches for each modality, each consisting of specialized neural networks tailored to the data type. For instance, Convolutional Neural Networks are used for image data, Recurrent Neural Networks or Transformers for text, and CNNs or specialized audio networks like Spectrogram-based models for audio. These branches extract high-level features from their respective modalities, capturing the essential information needed for prediction. The fusion of these modalities is a critical step, often achieved through techniques like concatenation, attention mechanisms, or more sophisticated methods such as tensor fusion or cross-modal transformers. Concatenation involves merging the feature vectors from each modality into a single vector, which is then fed into a fully connected network for final prediction. Attention mechanisms, on the other hand, dynamically weigh the importance of each modality based on the context, allowing the model to focus on the most relevant features. Cross-modal transformers can further enhance fusion by enabling interactions between modalities at multiple layers, capturing complex dependencies. The training process involves feeding the multimodal data into the model and optimizing it using a suitable loss function, such as cross-entropy for classification tasks or mean squared error for regression. The model is trained on a dataset containing labeled examples of multimodal inputs and their corresponding codewords or labels. Data augmentation techniques, such as random cropping for images or time-stretching for audio, can be applied to increase dataset diversity and improve generalization. Evaluation is performed using metrics like accuracy, precision, recall, and F1-score, depending on the nature of the prediction task. The model's performance is validated on a separate test set to ensure it generalizes well to unseen data. Additionally, visualization techniques, such as attention maps or feature embeddings, can provide insights into how the model is leveraging each modality. Finally, the program is deployed with an inference pipeline that preprocesses input data, feeds it into the trained model, and outputs the predicted codewords. This pipeline can be integrated into larger systems or made accessible via a user interface, enabling real-time predictions in practical applications.

3.3 The final combination of two models

It creates a powerful pipeline that not only improves the quality of low-light images but also leverages enhanced images alongside other data modalities for accurate codeword prediction. This integrated approach is particularly useful in applications like autonomous driving, surveillance, and multimedia analysis, where both image quality and multimodal context are critical for decision-making. The design begins with the low-light image enhancement module, which processes input images captured under poor lighting conditions. This module employs a deep learning model, such as a Convolutional Neural Network or a Generative Adversarial Network , trained to enhance low-light images by increasing brightness, reducing noise, and preserving details. The enhanced images are then passed to the multimodal fusion module, where they are combined with other data modalities, such as text, audio, or sensor data, to form a comprehensive input for codeword prediction. The multimodal fusion module is designed to handle diverse data types, each processed by specialized neural networks. For instance, CNNs are used for image data, RNNs or Transformers for text, and Spectrogram-based models for audio. These networks extract high-level features from their respective modalities, which are then fused using techniques like concatenation, attention mechanisms, or cross-modal transformers. The fusion process ensures that the model can effectively integrate information from all modalities, capturing complex dependencies and enhancing prediction accuracy. The combined program is trained end-to-end, with the low-light image enhancement module and the multimodal fusion module optimized jointly. The training dataset consists of low-light images, their corresponding enhanced versions, and additional modalities, along with the target codewords or labels. A suitable loss function, such as cross-entropy for classification tasks or mean squared error for regression, is used to guide the training process. Data augmentation techniques, such as random cropping for images and time-stretching for audio, are applied to improve generalization. During inference, the program first enhances low-light images using the image enhancement module. The enhanced images, along with other modalities, are then fed into the multimodal fusion module to predict the codewords. The program outputs the predicted codewords, which can be used for decision-making in various applications. For example, in autonomous driving, the enhanced images and fused multimodal data can help the vehicle navigate safely in low-light conditions.

3.4 The performance evaluation of the design

Evaluating the performance of a low-light image enhancement program and a multimodal fusion codeword prediction program requires a comprehensive set of metrics tailored to the specific tasks and objectives of each module. These metrics ensure that the programs not only meet technical benchmarks but also deliver practical value in real-world applications. For the low-light image enhancement program, the primary goal is to improve the visual quality of images captured under poor lighting conditions. Key metrics include Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise, providing a quantitative assessment of image quality improvement. Higher PSNR values indicate better enhancement. SSIM, on the other hand, evaluates the perceptual quality by comparing luminance, contrast, and structure between the enhanced and reference images. An SSIM value close to 1 indicates high similarity to the reference image. Additionally, Mean Squared Error (MSE) can be used to measure pixel-wise differences, while visual inspection remains crucial for subjective quality assessment. In the case of the multimodal fusion codeword prediction program, the focus shifts to the accuracy and reliability of predicting codewords or labels based on integrated multimodal data. Common metrics include accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly predicted codewords, while precision and recall evaluate the model's ability to identify relevant instances and avoid false positives and negatives, respectively. The F1-score, the harmonic mean of precision and recall, provides a balanced measure of the model's performance. For regression tasks, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used to quantify prediction errors. Additionally, confusion matrices and Receiver Operating Characteristic (ROC) curves can provide deeper insights into classification performance. When combining these two programs, it is essential to evaluate both the enhancement quality and the prediction accuracy. For instance, in an autonomous driving scenario, the enhanced images should not only be visually superior but also contribute to accurate object detection and navigation decisions. Therefore, the combined evaluation might include metrics like detection accuracy, mean Average Precision (mAP) for object detection, and task-specific performance indicators such as navigation success rate.

4. The summarization and prospect of the design

4.1 Application in vehicle networking communication

In the context of vehicular communication systems, the integration of low-light image enhancement and multimodal fusion codeword prediction programs offers a transformative approach to improving safety, efficiency, and user experience. These programs work in tandem to address the challenges posed by low-visibility conditions and the need for accurate, real-time decision-making based on diverse data sources. The process begins with the low-light image enhancement program, which is crucial for vehicles operating in environments with poor lighting, such as nighttime or tunnels. Cameras mounted on the vehicle capture low-light images, which are then processed by the enhancement module. Using deep learning models like CNNs or GANs, the program enhances these images by increasing brightness, reducing noise, and preserving critical details. The enhanced images provide clearer visual input for subsequent processing, enabling better detection of obstacles, road signs, and other vehicles. Once the images are enhanced, they are fed into the multimodal fusion codeword prediction program alongside other data modalities, such as LiDAR, radar, GPS, and vehicle-to-everything (V2X) communication signals. Each modality is processed by specialized neural networks—CNNs for images, RNNs or Transformers for text-based data, and spectrogram-based models for audio. The program then fuses these modalities using techniques like concatenation, attention mechanisms, or cross-modal transformers, capturing complex dependencies and ensuring a comprehensive understanding of the environment. The fused data is used to predict codewords or labels that represent critical information, such as the presence of pedestrians, the state of traffic lights, or the identification of road hazards. These predictions are crucial for real-time decision-making in autonomous driving systems, enabling the vehicle to navigate safely and efficiently. For instance, if the system detects a pedestrian crossing the road in low-light conditions, it can immediately alert the driver or initiate an emergency stop. Throughout this process, the programs operate in a continuous feedback loop. The low-light image enhancement module ensures that visual data is always of high quality, even in challenging lighting conditions, while the multimodal fusion codeword prediction module leverages this enhanced data alongside other inputs to make accurate and reliable predictions. This synergy is particularly valuable in dynamic and unpredictable environments, where quick and precise responses are essential.

4.2 Inspiration in developed beam training scheme

The advancements in low-light image enhancement and multimodal fusion codeword prediction programs offer valuable insights and inspiration for improving beam training in modern communication systems, particularly in vehicular networks and 5G/6G applications. These programs demonstrate the importance of adaptive, data-driven approaches and the integration of diverse information sources, which can be directly applied to enhance the efficiency and effectiveness of beam training. One key inspiration is the use of deep learning to address challenging environmental conditions. Just as low-light image enhancement programs leverage CNNs and GANs to improve image quality in poor lighting, beam training can benefit from similar models to optimize beamforming in complex and dynamic environments. For instance, deep learning can be used to predict optimal beam directions based on real-time channel conditions, reducing the time and computational resources required for traditional beam training methods. This is particularly relevant in high-mobility scenarios, such as vehicular communication, where rapid beam alignment is critical. Another important insight is the value of multimodal data fusion. Multimodal fusion codeword prediction programs excel at integrating diverse data sources, such as images, text, and audio, to make accurate predictions. Similarly, beam training can incorporate multiple data modalities, such as channel state information (CSI), GPS data, and environmental sensors, to enhance beam alignment. By fusing these modalities, beam training systems can achieve more robust and context-aware beamforming, adapting to changes in the environment and user mobility. The concept of continuous adaptation and feedback loops in these programs also provides a blueprint for beam training. Low-light image enhancement and multimodal fusion programs operate in real-time, continuously refining their outputs based on new inputs. Beam training systems can adopt a similar approach, using real-time feedback from the communication channel to dynamically adjust beam directions and maintain optimal performance. This is especially useful in scenarios with high user mobility or rapidly changing environmental conditions. Furthermore, the emphasis on efficiency and resource optimization in these programs can inspire improvements in beam training. Low-light image enhancement programs focus on delivering high-quality results with minimal computational overhead, while multimodal fusion programs prioritize accurate predictions using efficient fusion techniques. Beam training systems can similarly benefit from lightweight, efficient algorithms that reduce training overhead and improve scalability, particularly in large-scale networks with many users and antennas. the innovations provide valuable lessons for advancing beam training in modern communication systems by adopting data-driven approaches, integrating diverse data sources, and emphasizing real-time adaptation and efficiency.

5. Conclusion

This paper presents a design approach for an improved beam training scheme, which includes separately designing a low brightness image enhancement module and a codeword prediction module, and ultimately combining them to form a complete beam training technique. The performance evaluation indicators of this design have also been provided, which have significant application significance in vehicle networking communication systems and provide new inspiration for beam training. If the design is reasonably realized, it will help to improve the safety and stability of the auto drive system.

References

[1] U. Demirhan and A. Alkhateeb, "Radar Aided 6G Beam Prediction: Deep Learning Algorithms and Real-World Demonstration," 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 2022, pp. 2655-2660

[2] M. Dias, A. Klautau, N. Gonzalez-Prelcic, and R. W. Heath, "Position and LIDAR-aided mmWave beam selection using deep learning," in Proc. IEEE Int. Workshop on Signal Processing Adv. in Wireless Commun. (SPAWC), Cannes, France, 2019, pp. 1-5

[3] A. Klautau, N. Gonzalez-Prelcic, and R. W. Heath, "LIDAR data for deep learning-based mmWave beam-selection," IEEE Wireless Commun. Lett, vol. 8, no. 3, pp. 909-912, 2019

[4] M. Alrabeiah, A. Hredzak and A. Alkhateeb, "Millimeter Wave Base Stations with Cameras: Vision-Aided Beam and Blockage Prediction," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 2020, pp. 1-5

[5] G. Charan, T. Osman, A. Hredzak, N. Thawdar and A. Alkhateeb, "Vision-Position Multi-Modal Beam Prediction Using Real Millimeter Wave Datasets," 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 2022, pp. 2727-2731