

Article

SDDA-DeepSIC: A Spatial-Dimming-Aware Dual-Attention Deep Unfolding Network for Multi-User MIMO Visible Light Communications

Jianing Wang ^{1,2}  and Guifen Chen ^{1,*}¹ School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China² School of Electronic Information Engineering, Changchun University, Changchun 130022, China

* Correspondence: cgfchangchunligong@163.com

Abstract

Spatial dimming in multi-user, multiple-input, multiple-output visible light communication (MIMO-VLC) systems changes the LED activation pattern and effective channel structure, thereby affecting inter-user interference and posing additional challenges to multi-user detection. To address this issue, this paper develops a spatial-dimming- and inter-user-interference-aware DeepSIC detection framework. First, a physically based MIMO-VLC model is established using the Lambertian line-of-sight channel, and a channel-aware LED selection strategy is employed to preserve effective channel separability under the spatial-dimming constraint. Spatial dimming attention (SDA) and inter-user interference attention (IIA) are then introduced to characterize LED-dependent spatial features and target-dependent heterogeneous interference relationships, respectively. An adaptive fusion gate (AFG) is further employed to dynamically integrate the two feature branches. In addition, a BER-oriented optimization strategy is adopted to train the unfolded DeepSIC detector toward the final binary detection objective. Numerical simulations demonstrate that the proposed framework effectively characterizes multi-user interference under spatial dimming and provides favorable detection performance in the low-to-moderate SNR regime. More importantly, measured module-level ablation experiments identify IIA as the primary contributor to the detection improvement. Compared with the original DeepSIC, introducing IIA alone reduces the BER by 9.88% at 15 dB and 34.92% at 25 dB. Attention visualization further shows that IIA learns differentiated interference weights for different target users. Meanwhile, the combined SDA, IIA, and AFG results reveal a nontrivial coupling between spatial-dimming- and interference-aware representations, with adaptive fusion exhibiting SNR-dependent behavior. These results highlight the importance of jointly considering spatial-dimming-induced channel variations and heterogeneous inter-user interference in learning-based multi-user detection for MIMO-VLC systems and provide insights into the development of physically constrained spatial attention and adaptive interference-fusion mechanisms.



Received: 22 September 2026

Revised: 7 October 2026

Accepted: 12 October 2026

Published: 22 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: visible light communication; MIMO; spatial dimming; multi-user interference elimination; DeepSIC; dual attention mechanism

1. Introduction

Visible light communication (VLC), which utilizes the visible light spectrum for wireless information transmission, has gradually become one of the important candidate tech-

nologies for sixth-generation mobile communication (6G) [1–4]. VLC uses LED lighting equipment as the information transmitter to achieve the integration of lighting and communication through Intensity Modulation and Direct Detection (IM/DD). Compared with traditional RF communication, VLC has the advantages of rich spectrum resources, no spectrum authorization, strong electromagnetic interference resistance, low energy consumption, and high security. It has broad application prospects in indoor high-speed communication, smart home, industrial Internet, Internet of Vehicles and other fields. In order to further enhance the communication capacity of VLC systems, multiple-input multiple-output (MIMO) technology has been widely applied in VLC systems [5–8]. MIMO-VLC can fully utilize spatial multiplexing gain and achieve high-speed, multi-user concurrent communication by forming a transmitting array with multiple LEDs and a receiving array with multiple photo-detectors (PDs). However, as the number of LEDs and users continues to increase, there is strong spatial overlap between different emission beams, resulting in severe spatial crosstalk and multi-user interference (MUI) between users, leading to a significant increase in system error rate and becoming a key factor restricting the performance improvement of MIMO-VLC. On the other hand, actual lighting systems not only require meeting communication needs, but also must meet lighting requirements in different scenarios. Therefore, spatial dimming technology has gradually become an important research direction for MIMO-VLC. Spatial dimming achieves lighting control at different brightness levels by dynamically controlling the activation state and emission power of LED arrays. However, changes in LED activation modes and optical power distribution can further alter the optical channel characteristics, resulting in dynamic spatial interference relationships between different users. Traditional detection algorithms are difficult to effectively adapt to this complex environment.

At present, multi-user interference cancellation mainly adopts methods such as zero forcing (ZF) [9,10], minimum mean square error (MMSE) [11,12], serial interference cancellation (SIC) [13,14], and maximum likelihood detection (MLD) [15–18]. For example, Feng R H proposed the use of the Non-Orthogonal Multiple Access (NOMA) method in his research to design a precoding algorithm for signal-dependent noise (SDN-FSO) in free space optical channels. The study showed that it converges faster under imperfect continuous interference cancellation (SIC) and achieves significant SINR improvement [19]. Feng R developed a block coordinate descent method that includes convex approximation, greedy heuristic, and meta-heuristic search, which can significantly improve confidentiality under different conditions [20]. Among them, SIC can effectively reduce multi-user interference by eliminating detected user signals layer by layer, making it an important method for MIMO detection. However, traditional SIC heavily relies on accurate channel state information (CSI) and fixed detection order, which can easily lead to error propagation and significant degradation of detection performance in the case of dynamic channel changes caused by spatial dimming. In recent years, deep-learning technology has made rapid progress in the field of communication. DeepSIC, as a typical deep unfolding detection method, maps each iteration step of the SIC algorithm into a neural network layer, achieving an organic combination of model-driven and data-driven approaches, and achieving excellent performance in wireless MIMO detection [21–25]. The existing DeepSIC mainly targets traditional RF communication designs, and its network inputs usually only include received signals and CSI, without fully considering the effects of LED activation mode changes under spatial dimming conditions and dynamic evolution of multi-user spatial interference. At the same time, users and LED features are usually processed with unified weights in the network, which makes it difficult to fully explore the importance differences between different LEDs and users, limiting further improvement in detection performance. Therefore, how to fully utilize spatial dimming information, dynamically model the im-

portance of LED and user interference relationships, and integrate them into DeepSIC networks to achieve high-precision and multi-user interference cancellation in complex spatial dimming environments has become an important scientific problem that urgently needs to be solved in the current MIMO visible light communication field.

Based on this, this study proposes a spatial-dimming-aware dual-attention DeepSIC (SDDA DeepSIC) multi-user detection method, which uses a spatial dimming perception dual attention module to jointly model the relationship between LED spatial features and user interference, improving the multi-user detection performance in complex spatial dimming environments. In response to the lack of spatial dimming modeling in traditional DeepSIC, the LED spatial activation mode, optical power allocation, and user interference relationship are uniformly introduced into the deep unfolding detection framework to establish a correlation mechanism between spatial dimming and multi-user interference, providing a new theoretical basis for intelligent detection of spatial dimming MIMO-VLC. We design spatial dimming attention and inter-user interference attention with communication physics significance, break through the limitations of traditional deep learning that relies solely on data-driven approaches, achieve deep fusion of communication physics priors and deep-learning models, and improve the interpretability and generalization ability of the models.

2. Theoretical Foundation

2.1. MIMO Visible Light Communication (MIMO-VLC)

Visible light communication (VLC) is a wireless communication technology that utilizes flickering signals emitted by objects such as fluorescent lamps or light-emitting diodes (LEDs) to transmit information. The frequency of visible light ranges from 400 THz (780 nm wavelength) to 800 THz (375 nm wavelength) [26–30]. VLC systems typically use Intensity Modulation and Direct Detection (IM/DD), where the transmitting end modulates the LED emission intensity and the receiving end uses a photo-detector (PD) to detect the received optical power and complete signal recovery. In order to further improve system capacity, MIMO technology is introduced, which uses multiple LEDs to form a transmitting array at the transmitting end and multiple PDs to form a receiving array at the receiving end, achieving spatial multiplexing.

The typical MIMO-VLC model can be represented as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where $\mathbf{y}$, $\mathbf{H}$, $\mathbf{x}$, $\mathbf{n}$ are the received signal, the optical channel matrix, the transmitted signal, and Gaussian noise, respectively.

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & \dots & h_{mn} \end{bmatrix} \quad (2)$$

Each element represents the LOS optical gain between $LED_j \rightarrow PD_i$.

VLC mainly uses line-of-sight (LOS) propagation. The Lambertian model is

$$h_{ij} = \frac{(m+1)A}{2\pi d_{ij}^2} \cos^m(\phi) T_s(\psi) g(\psi) \cos(\psi) \quad (3)$$

where A , d_{ij} , m , ϕ , ψ refer to PD area, propagation distance, Lambert order, transmission angle, and reception angle, respectively. PD area refers to the effective photosensitive area of a photo-detector.

From the above formula, it can be seen that the Lambertian model determines the relationship between LED position, received power, and user interference, where the optical channel matrix $\mathbf{H}$ implicitly learns that the LED directly affects the user interference relationship.

2.2. Multi-User Interference

In multi-user MIMO visible light communication (MIMO-VLC) systems, multiple users typically share the same LED array and optical transmission resources [31–34]. Due to the overlap of multiple LED emission beams in space, the light signals sent by different users will be simultaneously superimposed at the receiving end. Therefore, while receiving their own effective signal, the target user is inevitably affected by interference from other user signals, resulting in multi-user interference (MUI). With the increasing number of system users and spatial reuse scale, multi-user interference has become a key factor restricting the detection performance and system capacity improvement of MIMO-VLC systems.

If the system serves K users simultaneously, the received signal can be represented as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (4)$$

where $\mathbf{h}_k$ refers to the optical channel matrix, symbol vector for user transmission, and additive Gaussian noise. The equation can be rewritten as

$$\mathbf{y} = \sum_{k=1}^K \mathbf{h}_k x_k + \mathbf{n} \quad (5)$$

where $\mathbf{h}_u$ represents the channel vector corresponding to the k th user. For the target user $\mathbf{h}_j$, the received signal can be further represented as

$$\mathbf{y} = \mathbf{h}_u x_u + \sum_{j \neq u} \mathbf{h}_j x_j + \mathbf{n} \quad (6)$$

Here, the first term represents the useful signal of the target user, while the second term represents multi-user interference from other users. Therefore, the essence of multi-user detection is to accurately separate target user signals in complex interference environments and effectively suppress interference generated by other users. The severity of multi-user interference mainly depends on the channel correlation between users. When different users have similar spatial positions or optical propagation paths, their channel vectors usually have high correlation, making it difficult for the receiving end to accurately distinguish the signals of each user, resulting in serious multi-user interference. The channel correlation of user i, j can be expressed as

$$\rho_{i,j} = \frac{|\mathbf{h}_i^T \mathbf{h}_j|}{\|\mathbf{h}_i\|_2 \|\mathbf{h}_j\|_2} \quad (7)$$

The closer $\rho_{i,j}$ is to 1, the more similar the two user channels are, and the more severe the mutual interference is; the closer $\rho_{i,j}$ is to 0, the more orthogonal the two user channels are, making it easier for the receiving end to achieve signal separation. To further describe the interference relationship of the entire system, the Gram matrix is usually used to represent the spatial coupling relationship between users, that is

$$\mathbf{G} = \mathbf{H}^T \mathbf{H} \quad (8)$$

Diagonal elements represent the channel energy corresponding to each user, rather than diagonal element $G_{i,j} = h_i^T h_j, i \neq j$, reflecting the degree of channel correlation between users i, j . The larger the non-diagonal elements, the stronger the spatial coupling between corresponding users, and the more severe the multi-user interference generated. Therefore, the Gram matrix can characterize the overall interference topology between users in the system, providing important physical prior information for multi-user detection algorithms.

Compared with traditional RF communication, multi-user MIMO-VLC systems are also affected by spatial dimming [35,36]. Due to the fact that LEDs not only perform communication functions, but also need to meet lighting requirements, their activation state and emitted light power will dynamically adjust with the dimming level, resulting in continuous changes in the effective optical channel of the system. Therefore, the interference relationship between users is not fixed and unchanging, but changes in real time with the spatial dimming mode. This dynamic change further increases the complexity of multi-user detection, making it difficult for traditional detection algorithms based on fixed channel assumptions to maintain stable performance.

2.3. Deep Successive Interference Cancellation

Model-driven deep learning has gradually become an important research direction in intelligent communication. Unlike traditional data-driven neural networks, model-driven methods fully utilize the existing mathematical models of communication systems, unfolding classical iterative algorithms into neural networks with fixed layers. While preserving the physical interpretability of the algorithm, the detection performance is further improved through network parameter learning [37]. DeepSIC (Deep Successive Interference Cancellation) is based on the idea that a deep unfolding detection method is proposed, which maps each iteration of the traditional SIC algorithm to an unfolding layer in the network, replaces the traditional detection function with a neural network, and achieves effective fusion of model-driven and data-driven methods.

If the estimated user symbol at the l iteration is $\hat{\mathbf{x}}^{(l)}$, then the traditional SIC can be represented as

$$\hat{\mathbf{x}}^{(l+1)} = f(\mathbf{y}, \mathbf{H}, \hat{\mathbf{x}}^{(l)}) \quad (9)$$

where $f(\cdot)$ represents a serial interference cancellation process, including residual signal calculation, user detection, and interference cancellation operations. DeepSIC uses trainable neural networks to replace traditional detection functions, and represents the above process as

$$\hat{\mathbf{x}}^{(l+1)} = f_{\theta_l}(\mathbf{y}, \mathbf{H}, \hat{\mathbf{x}}^{(l)}) \quad (10)$$

$f_{\theta_l}(\cdot)$ represents the neural network corresponding to the l -th expansion layer, and θ_l is the learnable parameter for that layer. After multiple iterations, the user symbol estimation is finally obtained:

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(L)} \quad (11)$$

Among them are the layers of network expansion.

Unlike traditional deep neural networks, the network structure of DeepSIC is directly derived from the SIC algorithm, with each layer corresponding to a physically significant interference cancellation process, thus having good interpretability. Meanwhile, due to each layer of the network only learning from the detection process in traditional algorithms, while retaining the interference reconstruction and iterative update mechanism, the required number of training parameters is much smaller than that of pure data-driven

neural networks, which results in faster training speed and better generalization ability. In addition, DeepSIC can fully utilize the prior knowledge of communication systems and still have good robustness in channel estimation errors, model mismatches, and complex interference environments. Therefore, in recent years, it has become an important representative method for model-driven intelligent detection.

Although DeepSIC can effectively alleviate the error propagation problem of traditional SIC, it still has certain limitations. Firstly, existing DeepSIC typically only utilizes received signals and channel state information as network inputs, without fully considering other physically significant prior information in communication systems. Secondly, in the process of feature learning, DeepSIC usually adopts a unified feature representation method, assigning equal importance to different users and spatial features, lacking the ability to adaptively model key interference sources. For MIMO visible light communication systems, spatial dimming dynamically changes the LED activation mode and optical power allocation, causing the interference relationship between optical channels and users to constantly change. The interference contribution generated by different users has obvious dynamics and non-uniformity. Traditional DeepSIC cannot fully utilize this spatial dimming information, so there is still room for further improvement in its detection performance in complex dimming environments.

3. SDDA DeepSIC System Architecture and Attention Detection Method

3.1. SDDA DeepSIC Overall Architecture

This paper proposes a spatial-dimming-aware dual-attention deep serial interference cancellation network called SDDA DeepSIC (Spatial-Dimming-aware Dual-Attention DeepSIC) to address the dynamic changes in LED activation modes, high correlation between user channels, and severe error propagation in traditional SIC in multi-user MIMO visible light communication systems under spatial dimming conditions. This method uses DeepSIC as the model-driven skeleton and integrates spatial dimming state, LED emission power, optical channel state, and inter-user interference topology into the iterative detection process. It also utilizes a dual attention mechanism to adaptively weight key light sources and main interfering users.

As shown in Figure 1, SDDA DeepSIC mainly consists of the following four parts: multi-user MIMO-VLC system and input information construction, spatial dimming perception dual attention module SDDA, DeepSIC deep unfolding detection network, and joint loss function for communication detection and optical constraints.

The overall mapping relationship can be expressed as

$$\hat{\mathbf{x}} = \mathcal{F}_{\Theta}(\mathbf{y}, \mathbf{H}, \mathbf{s}, \mathbf{p}, \mathbf{U}) \quad (12)$$

where $\mathbf{y}$ is the received signal, $\mathbf{H}$ is the optical channel matrix, $\mathbf{s}$ is the LED spatial dimming state, $\mathbf{p}$ is the LED optical power allocation vector, $\mathbf{U}$ represents user position or user topology information, Θ is all learnable parameters of the network, and $\hat{\mathbf{x}}$ is the final detected multi-user symbol.

Unlike traditional DeepSIC, which only uses received signals and user soft estimation for iterative detection, the proposed method explicitly models two types of complementary information at each unfolding stage: one is the spatial dimming feature determined by the LED activation state, transmission power, and optical channel; the other type is the interference characteristics between users determined by the correlation of user channels and the current soft detection results. Through adaptive fusion of the two types of features, the network can dynamically adjust the detection strategy according to the current dimming mode and interference state.

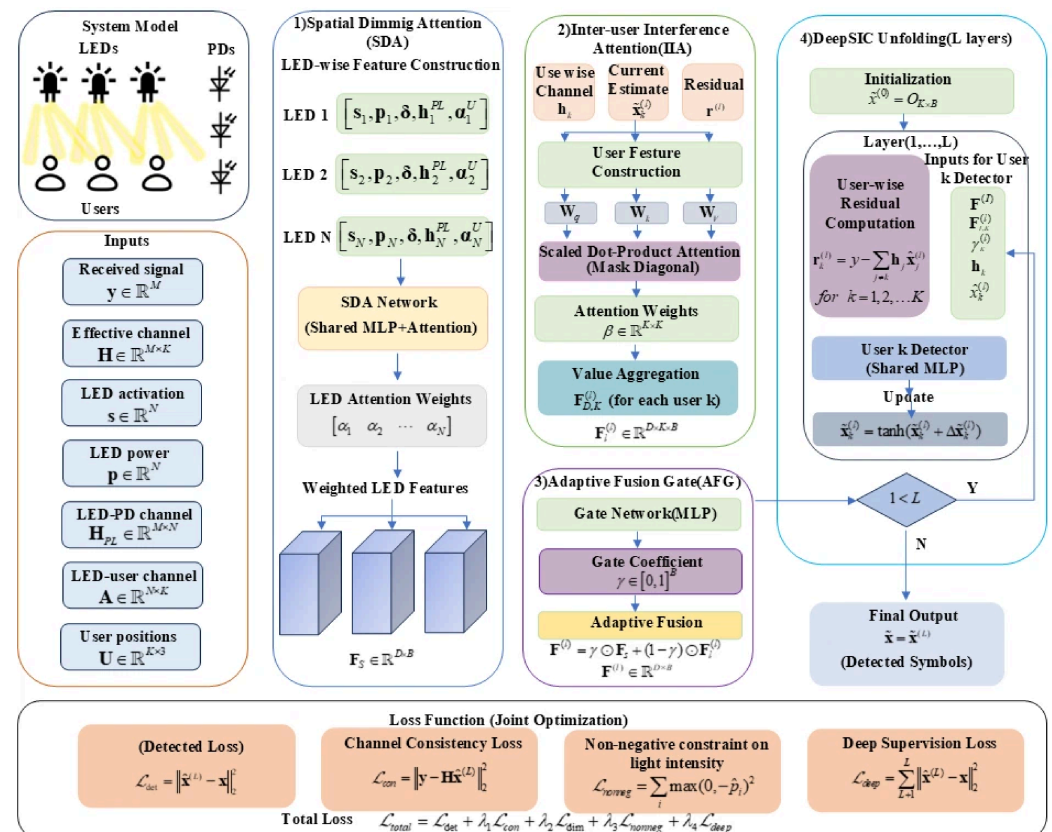


Figure 1. SDDA DeepSIC module schematic diagram.

3.2. Spatial Dimming MIMO-VLC System Model

The system includes N_{LED} LED light sources, K concurrent users, and M effective receiving dimensions. Multi-user signal reception can be represented as

$$\mathbf{y} = \mathbf{H}_{eff} \mathbf{x} + \mathbf{n} \quad (13)$$

where

$$\mathbf{y} \in \mathbb{R}^{M \times 1}, \mathbf{x} \in \mathbb{R}^{K \times 1} \quad (14)$$

representing received signals and multi-user transmission symbols, respectively, $\mathbf{n}$ is additive noise, and $\mathbf{H}_{eff} \in \mathbb{R}^{M \times K}$ is the effective optical channel under spatial dimming conditions.

The spatial dimming status of LED is represented as

$$\mathbf{s} = [s_1, s_2, \dots, s_{N_{LED}}]^T \quad (15)$$

$s_i \in \{0, 1\}$ represents the activation state of the i -th LED. The LED light power allocation vector is represented as

$$\mathbf{p} = [p_1, p_2, \dots, p_{N_{LED}}]^T, p_i \geq 0 \quad (16)$$

The spatial dimming state and optical power can be represented together as

$$\mathbf{d} = \mathbf{s} \odot \mathbf{p} \quad (17)$$

where $\odot$ represents the element-wise product. Spatial dimming changes the effective light source set and its power distribution, further altering the user's equivalent channel. Therefore, the effective channel can be abstractly represented as

$$\mathbf{H}_{eff} = \mathcal{H}(\mathbf{H}_0, \mathbf{d}, \mathbf{U}) \quad (18)$$

where $\mathbf{H}_0$ represents the basic optical channel when no spatial dimming is applied, and $\mathbf{U}$ represents the user's position, posture, or topological relationship.

For the target user k , the received signal can be written as

$$\mathbf{y} = \mathbf{h}_k x_k + \sum_{j \neq k} \mathbf{h}_j x_j + \mathbf{n} \quad (19)$$

where $\mathbf{h}_k x_k$ is the useful signal of the target user, and the second term is the multi-user interference generated by other users. Due to the simultaneous changes in $\mathbf{h}_k$ and $\mathbf{h}_j$ caused by spatial dimming, the interference relationship between users exhibits significant dynamism and non-uniformity. Therefore, fixed detection order or unified feature weights are difficult to adapt to the detection requirements under different dimming modes.

In a practical IM/DD transmitter, an appropriate DC operating point is required to ensure non-negative LED intensity. The present work does not explicitly optimize the DC bias or LED nonlinear transfer characteristics, since its focus is the multi-user detection mechanism under spatial dimming.

3.3. Input Information Construction and Shared Feature Embedding

To fully utilize the multi-source prior information in communication systems, this paper takes the received signal, channel matrix, spatial dimming state, optical power allocation, and user state as network inputs. The input for the first DeepSIC deployment phase can be represented as

$$\mathbf{z}^{(l)} = \left[\mathbf{y}; \text{vec}(\mathbf{H}_{eff}); \mathbf{s}; \mathbf{p}; \hat{\mathbf{x}}^{(l)} \right] \quad (20)$$

where $\text{vec}(\cdot)$ represents the matrix vectorization operation, and $\hat{\mathbf{x}}^{(l)}$ is the user soft estimation in the l stage.

$$\mathbf{F}_0^{(l)} = \phi(\text{Norm}(\mathbf{W}_e^{(l)} \mathbf{z}^{(l)} + \mathbf{b}_e^{(l)})) \quad (21)$$

Due to the different dimensions and statistical distributions of different inputs, they are first mapped to a unified feature space through linear projection, normalization, and nonlinear activation:

where $\mathbf{W}_e^{(l)}$ and $\mathbf{b}_e^{(l)}$ are trainable parameters; $\text{Norm}(\cdot)$ represents normalization operation; $\phi(\cdot)$ represents activation functions such as ReLU, GELU, or Tanh; and $\mathbf{F}_0^{(l)} \in \mathbb{R}^D$ represents shared feature representation.

Shared feature embedding has two functions: on the one hand, it eliminates the dimensional differences between the received signal, channel state, and dimming parameters; on the other hand, it provides a unified feature benchmark for subsequent SDA and IIA branches, avoiding the two attention branches learning in completely independent spaces, thereby improving the stability of dual branch fusion.

3.4. Spatial Dimming Perception Dual Attention Module

The SDDA module consists of four parts: spatial dimming attention, user interference attention, adaptive fusion gate, and residual enhancement. Its goal is to explore the

importance of LED and optical links in detection under dimming mode, as well as the interference sources of each target user.

3.4.1. Spatial Dimming Attention

Spatial dimming attention (SDA) is used to model the coupling relationship between LED activation mode, optical power distribution, and optical channel gain, and to assign adaptive weights to different LEDs. For the i th LED, its physical feature vector is constructed:

$$\mathbf{f}_i^{LED} = [s_i, p_i, \bar{h}_i, q_i]^T \quad (22)$$

where s_i is the active state, p_i is the emitted light power, $\bar{h}_i$ is the statistical representation of the channel gain related to the LED, and q_i can represent the local illumination, dimming level, or geometric relationship between the LED and the user. The composition of all LED features is as follows:

$$\mathbf{F}_{LED} = [\mathbf{f}_1^{LED}, \mathbf{f}_2^{LED}, \dots, \mathbf{f}_{N_{LED}}^{LED}]^T \quad (23)$$

To avoid obtaining unreasonably high weights by turning off the LED, an activation mask is introduced in the attention score. The attention score of the i th LED is

$$e_i^{SDA} = \mathbf{w}_S^T \phi(\mathbf{W}_s \mathbf{f}_i^{LED} + \mathbf{b}_s) + \mu \log(s_i + \varepsilon) \quad (24)$$

where μ is the activation mask strength, and ε is the minimum constant to prevent numerical overflow. After Softmax normalization, the LED attention weights are obtained:

$$\alpha_i = \frac{\exp(e_i^{SDA})}{\sum_{r=1}^{N_{LED}} \exp(e_r^{SDA})} \quad (25)$$

From this, spatial dimming perception features are obtained:

$$\mathbf{F}_s = \sum_{i=1}^{N_{LED}} \alpha_i \mathbf{W}_v^S \mathbf{f}_i^{LED} \quad (26)$$

The weight α_i represents the overall impact of the LED on useful signals and multi-user interference in the current dimming state. SDA provides an LED level feature selection mechanism related to dimming configuration for subsequent detection.

3.4.2. Interference of Attention Among Users

SDA extracts key spatial dimming information from the transmitting side, while inter-user interference attention (IIA) characterizes multi-user interference topology from the user side.

Let the effective channel matrix be

$$\mathbf{H}_{eff} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K] \quad (27)$$

The normalized channel correlation coefficient of user i, j can be defined as

$$\rho_{i,j} = \frac{|\mathbf{h}_i^T \mathbf{h}_j|}{\|\mathbf{h}_i\|_2 + \|\mathbf{h}_j\|_2 + \varepsilon} \quad (28)$$

The Gram matrix can be simultaneously constructed:

$$\mathbf{G} = \mathbf{H}_{eff}^T \mathbf{H}_{eff} \quad (29)$$

Diagonal elements represent the channel energy of each user, while non-diagonal elements reflect the channel coupling strength between users. To reduce the scale bias caused by user power differences, a normalized interference matrix can be further constructed:

$$\tilde{G}_{i,j} = \frac{|G_{ij}|}{\sqrt{G_{ii}G_{jj} + \varepsilon}} \quad (30)$$

For the k th user, their phased characteristics can be represented as

$$\mathbf{f}_k^{(l)} = [\mathbf{h}_k; \hat{x}_k^{(l)}; c_k^{(l)}; r_k^{(l)}] \quad (31)$$

where $\hat{x}_k^{(l)}$ is the current symbol soft estimation, $c_k^{(l)}$ is the estimation confidence, and $r_k^{(l)}$ is the residual statistic related to the user.

Query, Key, and Value are obtained through learnable mapping:

$$\mathbf{q}_k = \mathbf{W}_Q \mathbf{f}_k^{(l)}, \mathbf{k}_j = \mathbf{W}_K \mathbf{f}_j^{(l)}, \mathbf{v}_j = \mathbf{W}_V \mathbf{f}_j^{(l)} \quad (32)$$

The interference attention score of user j on target user k is defined as

$$e_{kj}^{\text{IIA}} = \frac{\mathbf{q}_k^T \mathbf{k}_j}{\sqrt{d_\alpha}} + \eta \tilde{G}_{k,j} + \zeta \omega_{kj}^{(l)} \quad (33)$$

where d_α is the attention feature dimension, η, ζ is the weight coefficient, and $\omega_{kj}^{(l)}$ represents the dynamic interference prior obtained based on the current soft estimation or residual signal.

To avoid interference attention from the target user, set $e_{kk}^{\text{IIA}} = -\infty$. Then, normalized interference weights are obtained through Softmax:

$$\beta_{kj} = \frac{\exp(e_{kj}^{\text{IIA}})}{\sum_{m \neq k} \exp(e_{km}^{\text{IIA}})} \quad (34)$$

The interference perception characteristics of the k th user are

$$\mathbf{F}_{I,k}^{(l)} = \sum_{j \neq k} \beta_{kj} \mathbf{v}_j \quad (35)$$

The composition of interference characteristics of all users is as follows:

$$\mathbf{F}_I^{(l)} = [\mathbf{F}_{I,1}^{(l)}, \dots, \mathbf{F}_{I,K}^{(l)}] \quad (36)$$

In the equations, IIA indicates that the network no longer considers all other users as equally important sources of interference, but dynamically assigns weights based on their channel correlation, symbol confidence, and residual contribution. For users with highly correlated channels, low estimation confidence, or large residual contributions, their attention weights are usually higher, allowing DeepSIC to prioritize and suppress key interference.

3.4.3. Adaptive Fusion Gate

SDA and IIA, respectively, describe the relationship between the spatial dimming structure on the transmitting side and the user interference on the receiving side. The importance of the two will vary with dimming level, user distribution, signal-to-noise ratio, and channel correlation, so they cannot be simply added in a fixed ratio.

Firstly, the two types of features are connected:

$$\mathbf{F}_I^{(l)} = [\mathbf{F}_{I,1}^{(l)}, \dots, \mathbf{F}_{I,K}^{(l)}] \quad (37)$$

Subsequently, fusion weights are generated through a gate controlled network:

$$\gamma^{(l)} = \sigma(\mathbf{W}_g^{(l)} \mathbf{F}_C^{(l)} + \mathbf{b}_g^{(l)}) \quad (38)$$

where $\sigma(\cdot)$ is the sigmoid function; therefore,

$$0 \leq \gamma_d^{(l)} \leq 1 \quad (39)$$

The final fusion feature is

$$\mathbf{F}_{fusion}^{(l)} = \gamma^{(l)} \odot \mathbf{F}_S^{(l)} + (1 - \gamma^{(l)}) \odot \mathbf{F}_I^{(l)} \quad (40)$$

When the system is at a lower dimming level and there are fewer effective LEDs, the LED activation mode has a more significant impact on the effective channel, and the gating network can improve the contribution of the SDA branch. When the user distribution is dense and the channel correlation is high, the network can enhance the IIA branch, thereby achieving dynamic adaptation to complex communication scenarios.

3.4.4. Residual Feature Enhancement

Attention weighting may suppress some original features with low weights but still have discriminative value. Therefore, this article adopts residual connections to preserve shared features:

$$\mathbf{F}_{out}^{(l)} = \text{Norm}(\mathbf{F}_0^{(l)} + \mathbf{W}_r^{(l)} \mathbf{F}_{fusion}^{(l)}) \quad (41)$$

Residual enhancement can reduce information loss caused by bias in attention weight estimation and improve gradient propagation in multi-layer unfolding networks. The $\mathbf{F}_{out}^{(l)}$ obtained from this includes original signal features, spatial dimming priors, and user interference priors, providing enhanced input for the current DeepSIC layer.

3.5. SDDA Enhanced DeepSIC Detection Method

3.5.1. Traditional SIC Detection Process

Traditional SIC first detects users in a preset order and sequentially eliminates the reconstructed signals of the detected users from the received signal. For target user k , the interference reconstruction of the l th iteration can be written as

$$\hat{\mathbf{I}}_k^{(l)} = \sum_{j \neq k} \mathbf{h}_j \hat{x}_j^{(l)} \quad (42)$$

The corresponding interference cancellation residual is

$$\mathbf{r}_k^{(l)} = \mathbf{y} - \hat{\mathbf{I}}_k^{(l)} \quad (43)$$

Subsequently, new symbol estimates are obtained using linear or nonlinear detectors:

$$\hat{x}_k^{(l+1)} = \mathcal{D}_k(\mathbf{r}_k^{(l)}, \mathbf{h}_k) \quad (44)$$

The main problem with traditional SIC is that once a user's estimation is incorrect, the error will be carried into interference reconstruction and continue to affect subsequent

user detection, resulting in error propagation. DeepSIC expands the iterative process of traditional SIC into a fixed-layer neural network. The l th expansion layer corresponds to a multi-user interference update, using trainable detectors instead of fixed symbol decision functions:

$$\hat{\mathbf{x}}^{(l+1)} = \mathcal{D}_{\theta_l}(\mathbf{y}, \mathbf{H}_{eff}, \hat{\mathbf{x}}^{(l)}) \quad (45)$$

Among them is the parameter of the θ_l layer. Unlike strict serial hard decision SIC, DeepSIC typically uses user soft estimation or category probability for interference reconstruction, thereby reducing the direct impact of preceding hard decision errors on subsequent detection.

3.5.2. Iterative Detection Driven by SDDA

In the proposed network, the expansion layer first calculates SDDA features based on the current symbol estimation:

$$\mathbf{F}_{out}^{(l)} = SDDA(\mathbf{y}, \mathbf{H}_{eff}, \mathbf{s}, \mathbf{p}, \mathbf{U}, \hat{\mathbf{x}}^{(l)}) \quad (46)$$

Subsequently, parallel interference reconstruction is performed using the current symbol estimation:

$$\hat{\mathbf{I}}^{(l)} = \mathbf{H}_{eff} \hat{\mathbf{x}}^{(l)} \quad (47)$$

The overall residual signal is

$$\mathbf{r}^{(l)} = \mathbf{y} - \hat{\mathbf{I}}^{(l)} \quad (48)$$

To avoid eliminating its own signal when updating the target user, user level residuals can be further constructed:

$$\mathbf{r}_k^{(l)} = \mathbf{y} - \sum_{j \neq k} \mathbf{h}_j \hat{\mathbf{x}}_j^{(l)} \quad (49)$$

The residual, current soft estimation, and SDDA enhanced features are sent together into the current expansion layer:

$$\Delta \mathbf{x}^{(l)} = f_{\theta_l}(\mathbf{r}^{(l)}, \hat{\mathbf{x}}^{(l)}, \mathbf{F}_{out}^{(l)}) \quad (50)$$

Subsequently, residual symbol updates were used:

$$\hat{\mathbf{x}}^{(l+1)} = \Pi_{\mathcal{X}}(\hat{\mathbf{x}}^{(l)} + \Delta \mathbf{x}^{(l)}) \quad (51)$$

where $\Pi_{\mathcal{X}}(\cdot)$ represents the feasible domain for mapping continuous outputs to modulation symbols. For BPSK soft detection, it is possible to use

$$\hat{\mathbf{x}}^{(l+1)} = \tanh(\hat{\mathbf{x}}^{(l)} + \Delta \mathbf{x}^{(l)}) \quad (52)$$

For Q-order discrete modulation, the class probability of each user can be output:

$$\mathbf{p}_k^{(l+1)} = \text{softmax}(\mathbf{o}_k^{(l)}) \quad (53)$$

This design forms a closed loop between attention weights and DeepSIC iteration states: the current user's soft estimation affects the modeling of interference relationships in IIA, and the updated interference attention further guides the next layer of symbol

detection. As the number of unfolding layers increases, the confidence level of erroneous users is gradually corrected, and the main sources of interference are dynamically updated, thus achieving iterative attention detection and interference elimination.

3.6. Physical Constraint Joint Loss Function

To simultaneously optimize multi-user detection performance, channel consistency, and spatial dimming constraints, a joint loss function can be defined:

$$\mathcal{L}_{total} = \mathcal{L}_{det} + \lambda_1 \mathcal{L}_{con} + \lambda_2 \mathcal{L}_{dim} + \lambda_3 \mathcal{L}_{nonneg} + \lambda_4 \mathcal{L}_{deep} \quad (54)$$

where $\mathcal{L}_{det}, \mathcal{L}_{con}, \mathcal{L}_{dim}, \mathcal{L}_{nonneg}, \mathcal{L}_{deep}$ refer to symbol detection loss, channel consistency loss, dimming loss, non-negative constraint loss of light intensity, and deep supervision loss, respectively.

With regard to symbol detection loss $\mathcal{L}_{det}$, for categorical symbol detection, cross-entropy loss is used:

$$\mathcal{L}_{det} = -\frac{1}{BK} \sum_{b=1}^B \sum_{k=1}^K \sum_{q=1}^Q t_{bkq} \log p_{bkq} \quad (55)$$

where B, t_{bkq}, p_{bkq} represent batch size, real label, and prediction probability, respectively.

In order to make the predicted symbols consistent with the actual received signal after physical channel reconstruction, channel consistency loss $\mathcal{L}_{con}$ is defined as

$$\mathcal{L}_{con} = \frac{1}{B} \sum_{b=1}^B \left\| \mathbf{y}_b - \mathbf{H}_{eff,b} \hat{\mathbf{x}}_b \right\|_2^2 \quad (56)$$

This can constrain the network output to conform to the MIMO-VLC receiver model, rather than just pursuing label fitting.

For dimming loss $\mathcal{L}_{dim}$, the target dimming ratio is first set to δ_{tar} , and the actual average dimming ratio and dimming loss are, respectively, set to

$$\delta = \frac{1}{N_{LED}} \sum_{i=1}^{N_{LED}} s_i \quad (57)$$

$$\mathcal{L}_{dim} = (\delta - \delta_{tar})^2 \quad (58)$$

It is worth noting that if the dimming matrix and power allocation are predetermined by the system rather than the network output, then this term should not be directly used as a learnable loss for the detection network but rather as a data validity constraint or for joint dimming detection optimization.

When the network predicts both light intensity and power variables simultaneously, the non-negative constraint $\mathcal{L}_{nonneg}$ on light intensity can be defined as

$$\mathcal{L}_{nonneg} = \sum_i \max(0, -\hat{p}_i)^2 \quad (59)$$

If the network only outputs bipolar baseband symbols without directly controlling the optical power, non-negative constraints should not be applied directly to the symbol output but should be applied to the LED emission signal or predicted power variable after DC bias.

Regarding deep supervision loss $\mathcal{L}_{deep}$, to improve the training of intermediate unfolding layers, auxiliary supervision can be applied to the output of each layer:

$$\mathcal{L}_{deep} = \sum_{l=1}^L \omega_l \mathcal{L}_{det}^{(l)} \quad (60)$$

where the later expansion layer is usually given greater weight, that is

$$\omega_1 \leq \omega_2 \leq \dots \leq \omega_L \quad (61)$$

Deep supervision can reduce the insufficient gradient in early layers and enable each unfolding layer to have effective interference correction capabilities.

3.7. Complete Testing Process

This paper proposes the SDDA DeepSIC detection algorithm to address the dynamic changes in LED activation status and complex interference coupling between users in multi-user MIMO-VLC systems under spatial dimming conditions. This algorithm embeds spatial dimming attention (SDA), inter-user interference attention (IIA), and adaptive fusion gate (AFG) into the DeepSIC iterative unfolding process, achieving multi-user signal detection through user level interference cancellation and layer by layer soft symbol updates. The specific process is as follows. The end-to-end detection process of SDDA DeepSIC is shown in Process 1:

Step 1: System Information Input and Symbol Initialization

Obtain received signal $\mathbf{y}$, optical channel matrix $\mathbf{H}_{eff}$, spatial dimming state $\mathbf{s}$, LED power allocation $\mathbf{p}$, and user topology $\mathbf{U}$

Step 2: Share communication feature encoding

$$\mathbf{F}_0^{(l)} = \phi(\text{Norm}(\mathbf{W}_e^{(l)} \mathbf{z}^{(l)} + \mathbf{b}_e^{(l)})) \quad \text{Formula ((13)–(19))}$$

Step 3: Build LED level spatial dimming features

For the i th LED, its activation state, emission power, system dimming ratio, LED-PD channel, and LED user spatial coupling information are jointly constructed into LED

level physical features: $\mathbf{f}_i^{LED} = [s_i, p_i, \bar{h}_i, q_i]^T$ Formula ((20)–(23))

Step 4: SDA Spatial Dimming Attention Calculation

$$e_i^{SDA} = \mathbf{w}_s^T \phi(\mathbf{W}_s \mathbf{f}_i^{LED} + \mathbf{b}_s) + \mu \log(s_i + \varepsilon) \quad \text{Formula ((24)–(26))}$$

Step 5: Calculate user level SIC Formula ((27)–(36))

Step 6: Interference Attention Calculation among IIA Users Formula ((37)–(40))

Step 7: AFG Dual Attention Feature Adaptive Fusion Formula (41)

Step 8: User level DeepSIC symbol detection and update Formula ((46)–(53))

Step 9: Iterative unfolding and final decision of multi-layer DeepSIC

$$\tilde{x}_k = \begin{cases} +1, \tilde{x}_k^{(L)} \geq 0 \\ -1, \tilde{x}_k^{(L)} < 0 \end{cases}$$

Step 10: Joint loss function and end-to-end parameter optimization,

During the training phase, end-to-end optimization is performed on the entire SDDA DeepSIC, and the total loss function is expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{det} + \lambda_1 \mathcal{L}_{con} + \lambda_2 \mathcal{L}_{dim} + \lambda_3 \mathcal{L}_{nonneg} + \lambda_4 \mathcal{L}_{deep} \quad \text{Formula (54)}$$

4. Experiment and Results Analysis

4.1. Experimental Setup and Dataset

To verify the interference cancellation capability of the proposed SDDA DeepSIC in spatially dimming multi-user MIMO visible light communication systems, this paper

constructs a multi-user MIMO-VLC simulation platform based on MATLAB (MATLAB-R2023b) and generates a synthetic dataset containing different user positions, spatial dimming modes, optical power allocation, and signal-to-noise ratio conditions. The experiment mainly examines the symbol detection performance, convergence characteristics, and generalization ability of the model under dynamic spatial dimming and multi-user channel-related conditions.

4.1.1. Multi-User MIMO-VLC Simulation Scenario

The simulation system consists of an LED emitting array, a multi-user optical channel, and a PD receiving array. Assuming that the sending end is configured with $N_{LED} = 8$ LED light sources, the system simultaneously serves $K = 4$ users, and the receiving end contains $M = 4$ valid PD receiving dimensions. LED light sources are distributed in a regular grid at the top of the normalized indoor area, and user positions are randomly generated within the effective indoor range. To avoid extreme channel gain caused by the user being located in the boundary area, the two-dimensional position coordinates of the user are uniformly sampled within the interval $[0.1, 0.9] \times [0.1, 0.9]$. For each sample, user positions are generated independently, allowing the dataset to cover different user distances, user aggregation levels, and channel correlation conditions. The user position matrix in the b th sample is represented as

$$U_b = [u_{1,b}, u_{2,b}, \dots, u_{K,b}]^T \in \mathbb{R}^{K \times 2} \quad (62)$$

where $u_{k,b}$ is the two-dimensional spatial coordinate of the k th user.

4.1.2. Spatial Dimming Mode Generation

In order to simulate the dynamic changes in lighting brightness, this paper adopts a spatial dimming method based on LED activation status. The spatial dimming vector is defined as $s_b = [s_{1,b}, s_{2,b}, \dots, s_{LED,b}]^T$, where $s_{i,b} \in \{0, 1\}$ represents the activation state of the i th LED in the b th sample. The target dimming ratio is set to $\delta_{tar} = 0.5$, which means that on average, about 50% of the LEDs are in an active state. To enhance data diversity, the actual number of activated LEDs in each sample fluctuates randomly around the target value and satisfies $1 \leq N_{active} \leq N_{LED}$. After determining the set of activated LEDs, a corresponding number of LEDs are randomly selected from all LEDs as effective light sources. This generation method can generate multiple dimming modes with different spatial distributions, enabling the model to learn the impact of LED activation position changes on effective channel and user interference.

The LED light power vector is represented as $p_b = [p_{1,b}, p_{2,b}, \dots, p_{LED,b}]^T$. To activate the LED, its power is randomly generated within a preset range. To turn the LED off, its power is set to 0. The power is specifically represented as $p_{i,b} = s_{i,b} \tilde{p}_{i,b}$, where $\tilde{p}_{i,b} \in [0.7, 1.3]$ is randomly sampled. Subsequently, the activation LED power is normalized to ensure that the average activation light power under different dimming modes is of a similar magnitude, avoiding strong bias on model training caused by the difference in total power.

4.1.3. Optical Channel Model and Signal Noise Model

Considering spatial dimming, LED and user geometric positions, and PD reception mixing effects, this paper represents the effective channel of the b th sample as

$$\mathbf{H}_{eff,b} = \mathbf{H}_{PD-LED,b} \mathbf{D}(s_b \odot p_b) \mathbf{A}_{LED-user,b} \quad (63)$$

where $\mathbf{H}_{PD-LED,b} \in \mathbb{R}^{M \times N_{LED}}$ represents the positive optical mixing matrix between the LED and PD receiving dimensions, $D(\cdot)$ represents a diagonal matrix composed of spatial dimming state and optical power, and $\mathbf{A}_{LED-user,b} \in \mathbb{R}^{N_{LED} \times K}$ represents the spatial coupling matrix between the LED and the user.

The coupling coefficient between the LED and the user is constructed based on their two-dimensional spatial distance:

$$\alpha_{ik,b} = \frac{1}{\varepsilon_d + \|l_i - u_{k,b}\|_2^2} \quad (64)$$

where l_i is the position of the i th LED, $u_{k,b}$ is the position of the k th user, and ε_d is a constant to prevent unstable values caused by too small a distance. Then, the coupling coefficients corresponding to each user are normalized:

$$\tilde{\alpha}_{ik,b} = \frac{\alpha_{ik,b}}{\sum_{r=1}^{N_{LED}} \alpha_{rk,b} + \varepsilon} \quad (65)$$

This model can reflect the distance difference between users and different LEDs and enable neighboring users to have more similar optical channels, thereby forming strong multi-user interference. To control the overall channel scale between different samples, the generated effective channel matrix is further normalized according to the Frobenius norm:

$$\tilde{\mathbf{H}}_{eff,b} = \frac{\mathbf{H}_{eff,b}}{\|\tilde{\mathbf{H}}_{eff,b}\|_F / \sqrt{MK} + \varepsilon} \quad (66)$$

It should be pointed out that this experiment uses synthetic optical channel data, mainly to verify the effectiveness of the proposed network structure and attention mechanism. In subsequent work, precise indoor VLC channels based on the Lambertian radiation model, non-line-of-sight reflection components, PD field of view angle, and LED nonlinearity can be further introduced to enhance the physical authenticity of the experimental environment.

The transmission symbol vector of the b -th sample is

$$\mathbf{x}_b = [x_{1,b}, x_{2,b}, \dots, x_{K,b}]^T \quad (67)$$

where $x_{k,b} \in \{-1, +1\}$.

The received signal is represented as

$$\mathbf{y}_b = \tilde{\mathbf{H}}_{eff,b} \mathbf{x}_b + \mathbf{n}_b \quad (68)$$

where $\mathbf{n}_b$ is additive Gaussian white noise. Noise power is calculated based on signal power and the specified signal-to-noise ratio:

$$\sigma_{n,b}^2 = \frac{P_{sig,b}}{10^{SNR_b/10}} \quad (69)$$

$$P_{sig,b} = \frac{1}{M} \|\tilde{\mathbf{H}}_{eff,b} \mathbf{x}_b\|_2^2 \quad (70)$$

The signal-to-noise ratio in the training and validation sets is randomly sampled within the range of 5–25 dB, enabling the network to learn robust symbol detection rules at different noise levels. During the testing phase, independent test data is generated at 5 dB intervals within the range of 0–30 dB to evaluate the BER performance of the model under different signal-to-noise ratio conditions.

Relationship to the physical VLC channel model: The above effective channel is a task-oriented simplification of the conventional Lambertian VLC channel model introduced in Section 2. Rather than reproducing all optical propagation parameters explicitly, the synthetic model retains the physical factors that are most relevant to the multi-user detection problem considered in this work. Specifically, the distance-based LED–user coupling preserves the spatial variation of channel gains caused by different LED–user geometries, while the spatial dimming and power matrices explicitly characterize LED activation states and transmit-power variations. The LED–PD mixing matrix further represents the spatial coupling between multiple optical transmit and receive dimensions. In contrast, detailed variations in irradiance and incidence angles and optical filter and concentrator gains are simplified or absorbed into the equivalent coupling coefficients under the assumed fixed indoor configuration. Higher-order reflections, blockage, multipath propagation, and device nonlinearities are not explicitly considered in the synthetic dataset. Frobenius normalization is applied to reduce the overall channel-scale variation and emphasize the relative spatial channel structure and inter-user interference patterns. This controlled formulation is sufficient for the primary objective of the present experiments, namely, evaluating the capability of SDDA-DeepSIC to exploit spatial dimming information and user-dependent interference relationships.

4.1.4. Dataset Partitioning and Parameters

Each data sample generated in this article includes $\{y_b, H_{eff,b}, s_b, p_b, U_b, x_b\}$, corresponding to the received signal, effective optical channel, spatial dimming state, LED power allocation, user location, and real transmission symbol. The training set contains 12,000 samples, with the SNR randomly sampled from 5 to 25 dB, and is used for network parameter optimization. The validation set contains 1500 independently generated samples within the same SNR range of 5–25 dB and is used for model selection and convergence monitoring. For performance evaluation, the test SNR is varied from 0 to 30 dB with an interval of 5 dB, resulting in seven SNR points, i.e., $\{0, 5, 10, 15, 20, 25, 30\}$ dB. At each SNR point, 3000 independent test samples are generated. Therefore, the complete test set contains 21,000 samples (7 SNR points $\times$ 3000 samples per SNR point). The test samples are used to evaluate the BER and generalization performance of the proposed detector. The user locations, spatial dimming configurations, LED power allocations, and noise realizations in the training, validation, and test sets are generated independently to prevent data leakage and to evaluate the generalization capability of the model to unseen spatial distributions and dimming configurations.

SDDA DeepSIC uses five DeepSIC expansion layers. The shared feature dimension is set to 64, and the projection dimension of user attention is set to 32. During the training process, the Adam optimizer is used, with an initial learning rate set to $\eta = 1 \times 10^{-3}$ and a batch size set to 64. The maximum number of training epochs is set to 30. The first-order and second-order moment decay coefficients of the Adam optimizer are set to 0.9 and 0.999, respectively.

The main parameters used for network training are: LED quantity $N_{LED} = 8$; user quantity $K = 4$; PD reception dimension $M = 4$; DeepSICC unfolding layer $L = 5$; shared feature dimension $D = 64$; attention dimension $d_a = 32$; batch size = 64; epochs = 30; the initial learning rate is GG, and the first-order and second-order decay coefficients of Adam are 0.9 and 0.999, respectively; the target dimming ratio is 0.5; the training SNR range is 5–25 dB; and the testing SNR range is 0–30 dB. To ensure the reproducibility of the training results, all experiments use fixed random seeds. During the model training process, the training loss and validation loss for each round are saved, and the convergence state of the model is monitored using the validation set performance.

Section 3.6 provides a general composite loss formulation for SDDA-DeepSIC, whereas only the loss components relevant to the specific prediction task are activated in the numerical experiments. In our experimental configuration, the network predicts only the transmitted user symbols. The spatial dimming matrix $\mathbf{S}$ and LED transmit power $\mathbf{P}$ are known system parameters and are directly provided to the network as physical inputs. Therefore, these quantities do not need to be estimated or reconstructed during training. In addition, their physical constraints are inherently satisfied by $\mathbf{S}_i \in \{0,1\}$ and $\mathbf{P}_i \geq 0$. Due to the current network only predicting user symbols, with spatial dimming matrix and LED power as known system inputs, the training loss mainly consists of symbol detection loss, channel consistency loss, and deep supervision loss, as shown in Section 3.6 above. In the experiment, $\lambda_{con} = 0.08, \lambda_{deep} = 0.20, \lambda_{att} = 1 \times 10^{-4}$ was set. Larger weights in the later layers can prompt the network to improve the symbol estimation accuracy layer by layer; the channel consistency constraint is used to ensure that the predicted symbols remain consistent with the received signal after effective channel reconstruction.

4.1.5. Physical Modeling Correction Modeling

From a practical communication perspective, the proposed spatial-dimming-aware effective channel construction aims to preserve sufficient spatial separability under the fixed dimming constraint, thereby enabling the multi-user MIMO-VLC system to approach or operate below the commonly adopted pre-FEC BER reference of 3.8×10^{-3} at practical SNR levels.

A physically based Lambertian LOS model is employed to construct the optical MIMO channel $\mathbf{H}_{LOS} \in \mathbb{R}^{K \times N}$, where each channel coefficient is determined by the LED-PD geometry and optical parameters. To impose spatial dimming, a binary LED activation vector $s \in \{0,1\}^{N_t}$ is introduced. The dimming ratio is strictly fixed at $\eta = 0.5$, such that $N_a = \eta N_t$ LEDs are active for every channel realization. Therefore, for the considered $N_t = 8$, exactly four LEDs are activated rather than allowing the number of active LEDs to vary randomly.

Since different LED activation patterns result in different effective MIMO channel structures, random LED activation may lead to poorly conditioned subchannels and degraded spatial separability. Therefore, channel-aware spatial dimming is adopted under the fixed dimming constraint. For each channel realization, all candidate LED subsets $\mathcal{A}$.

Satisfying $|\mathcal{A}| = N_a$ are considered, and the active subset is selected according to

$$\mathcal{A}^* = \arg \max_{\mathcal{A}: |\mathcal{A}|=N_a} \sigma_{\min}(\mathbf{H}_{LOS}(:, \mathcal{A})) \quad (71)$$

Maximizing the minimum singular value prevents severe degradation of the weakest spatial mode and thereby improves the spatial separability of the effective multi-user channel while maintaining exactly the same dimming ratio. The resulting received signal is expressed as

$$\mathbf{H}_{LOS} \mathbf{S} \mathbf{P} \mathbf{W}_{sel} \mathbf{x} + \mathbf{n} = \mathbf{H}_{eff} \mathbf{x} + \mathbf{n} \quad (72)$$

where $\mathbf{S} = \text{diag}(s)$, $\mathbf{P}$ denotes the LED power coefficients, and $\mathbf{W}_{sel}$ is the direct LED-stream selection matrix. This construction provides the effective channel used by the subsequent DeepSIC detector and attention modules.

4.1.6. Comparison Methods and Evaluation Indicators

In order to comprehensively evaluate the performance of the proposed SDDA DeepSIC algorithm, this paper conducts experimental evaluations from five aspects: detection performance, training convergence, model complexity, robustness, and module effectiveness. The specific evaluation indicators are as follows:

(1) Bit Error Rate (BER)

Bit Error Rate (BER) is the most important performance indicator in digital communication systems, used to measure the accuracy of the receiving end in recovering the transmitted bits. BER is defined as the proportion of erroneous bits to the total number of transmitted bits:

$$BER = \frac{N_{error}}{N_{total}} \quad (73)$$

where N_{error} is the number of bits for error detection and N_{total} is the total number of bits sent. The lower the BER, the better the system detection performance.

(2) Symbol Error Rate (SER)

Since the goal of multi-user detection is to recover the transmitted symbols, symbol error rate (SER) is further used to evaluate symbol level detection performance, which is defined as

$$SER = \frac{N_{symbol,error}}{N_{symbol,total}} \quad (74)$$

where $N_{symbol,error}$ is the number of symbols used to detect errors and $N_{symbol,total}$ is the total number of symbols sent.

The BPSK modulation used in this article has a consistent trend in SER and BER due to each symbol containing only one bit. For higher-order modulation methods such as PAM-4 or QAM, SER can more accurately reflect symbol detection performance.

(3) Detection Accuracy

Detection accuracy is used to measure the proportion of correctly detected symbols, defined as

$$Accuracy = \frac{N_{correct}}{N_{total}} \times 100\% \quad (75)$$

where $N_{correct}$ represents the number of correctly detected symbols and N_{total} represents the total number of detection symbols. The higher the accuracy, the more accurately the network can distinguish the symbols sent by different users.

(4) Mean Square Error (MSE)

To evaluate the difference between the network output and the actual transmitted symbols, mean square error (MSE) is used as the continuous value prediction error metric, which is defined as

$$MSE = \frac{1}{BK} \sum_{b=1}^B \|\hat{x}_b - x_b\|_2^2 \quad (76)$$

where B is the batch size, K is the number of users, $\hat{x}_b$ is the prediction symbol, and x_b is the real sending symbol.

The smaller the MSE, the closer the predicted result is to the actual transmitted symbol.

(5) Computational Complexity of the Model

In order to evaluate the actual deployment capability of the algorithm, this paper separately calculates the model parameter size (Number of Parameters), single sample inference time (Inference Time), and theoretical computational complexity.

The parameter count is defined as the total number of trainable parameters in the network:

$$N_{param} = \sum_i n_i \quad (77)$$

where n_i represents the number of trainable parameters in the i th layer. The smaller the number of parameters, the lower the model storage overhead.

The inference time represents the average time required for the model to complete the detection of a test sample:

$$T_{\text{inf}} = \frac{T_{\text{total}}}{N_{\text{sample}}} \quad (78)$$

where T_{total} is the total inference time and N_{sample} is the number of test samples. The shorter the inference time, the better the real-time detection capability of the algorithm.

The number of Floating Point Operations (FLOPs) in the forward propagation process of the network is further estimated, which is used to measure the computational complexity of the model. For SDDA DeepSIC, its overall complexity mainly consists of shared feature extraction, spatial dimming attention, user interference attention, and DeepSIC unfolding layers, which can be expressed as

$$\mathcal{O}(L(K^2d + Nd)) \quad (79)$$

where L is the number of unfolding layers, K is the number of users, N is the number of LEDs, and d is the feature dimension.

(6) Robustness

To verify the stability of the algorithm in complex communication environments, tests were conducted from two aspects: spatial dimming variation and channel estimation error. Different spatial dimming ratios were set $\delta \in \{0.25, 0.375, 0.50, 0.625, 0.75\}$, BER changes were calculated separately, and the algorithm's adaptability to different lighting brightness configurations was evaluated.

Varying degrees of estimation error were added to the channel matrix:

$$H' = H + \Delta H \quad (80)$$

$$\Delta H \sim \mathcal{N}(0, \sigma^2) \quad (81)$$

The BER changes under different error variances were statistically analyzed to evaluate the robustness of the algorithm to CSI errors.

4.2. Analysis of Experimental Results

4.2.1. Simulation Data Validation

(1) Basic data validation

We first verify the validity of the data selection. We define $\varepsilon = 10^{-6}$. $\varepsilon = 10^{-6}$ is only used to avoid numerical instability when the denominator approaches zero and is not an adjustable hyperparameter that affects model performance. A sensitivity analysis was conducted for $\varepsilon \in \{10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}\}$. Identical BER values were obtained at both 25 and 35 dB over the investigated range, indicating that the detector is insensitive to the numerical stabilization constant.

IIA query dimension: 32; IIA key dimension: 32; IIA value dimension: 32; AFG activation: sigmoid; residual enhancement activation: tanh; DeepSIC hidden activation: ReLU; DeepSIC output/update activation: tanh; weight initialization: Glorot; initial learning rate: 1×10^{-3} ; maximum epochs: 40; early-stopping patience: seven epochs; minimum validation improvement: 1×10^{-4} . The SDA module maps the LED-related spatial features to the shared $D = 64$ feature space. The IIA query, key, and value projections map the shared features from $D = 64$ to $D_q = 32$. The resulting interference-aware representation is subsequently mapped back to the shared feature space before fusion. AFG operates element-wise in the $D = 64$ shared feature space and therefore produces a 64-dimensional gating vector for each user.

(2) Effect of Unfolding Depth on Detection Performance

To investigate the influence of the unfolding depth, the number of DeepSIC stages was varied as $L = \{1, 2, 3, 5, 7\}$, while all other model and training configurations were kept unchanged. As shown in Table 1, increasing L from 1 to 3 substantially improves the detection performance, demonstrating the benefit of iterative interference cancellation. At 25 dB, the BER decreases from 2.345×10^{-2} for $L = 1$ to 4.40×10^{-3} for $L = 2$, and further to 2.35×10^{-3} for $L = 3$.

Table 1. Effect of the number of DeepSIC unfolding stages.

L	Parameters	BER = 15 dB	BER = 25 dB	BER = 30 dB	BER = 35 dB
1	21,090	3.125×10^{-2}	2.345×10^{-2}	2.325×10^{-2}	2.310×10^{-2}
2	25,891	1.075×10^{-2}	4.400×10^{-3}	4.100×10^{-3}	4.000×10^{-3}
3	30,692	7.250×10^{-3}	2.350×10^{-3}	2.050×10^{-3}	2.100×10^{-3}
5	40,294	6.900×10^{-3}	1.950×10^{-3}	2.000×10^{-3}	2.000×10^{-3}
7	49,896	8.350×10^{-3}	2.750×10^{-3}	2.600×10^{-3}	2.550×10^{-3}

Further increasing the depth to five stages reduces the BER to 1.95×10^{-3} . However, the incremental improvement becomes relatively small in the high-SNR regime; for example, the BER improvement from $L = 3$ to $L = 5$ at 30 dB is only approximately 2.44%. When the depth is further increased to $L = 7$, no additional gain is observed. Instead, the BER increases to 2.75×10^{-3} , 2.60×10^{-3} , and 2.55×10^{-3} at 25, 30, and 35 dB, respectively.

Meanwhile, the number of trainable parameters increases from 40,294 for $L = 5$ to 49,896 for $L = 7$, representing an approximately 23.8% increase in model size. These results suggest that the benefit of additional unfolding stages gradually saturates and that excessive depth does not necessarily lead to improved interference cancellation. Therefore, $L = 5$ is adopted as the default configuration because it provides the most favorable BER–complexity tradeoff among the investigated settings.

(3) Effect of Effective-Channel Normalization

To investigate the influence of channel-scale normalization, the proposed detector was additionally evaluated without the per-realization Frobenius normalization applied to the effective MIMO channel. In the baseline configuration, the effective channel is normalized according to

$$\tilde{\mathbf{H}}_{\text{eff}} = \frac{\mathbf{H}_{\text{eff}}}{\|\mathbf{H}_{\text{eff}}\|_F / \sqrt{MK} + \epsilon} \quad (82)$$

This operation reduces channel-scale variations and improves numerical conditioning during network optimization, but it also removes part of the absolute channel-gain information.

Table 2 compares the two configurations. The mean clean-signal power decreases from 3.9943 to 0.7838 when Frobenius normalization is removed, confirming that the normalization substantially changes the overall signal scale. Correspondingly, the BER increases from 7.75×10^{-3} to 9.20×10^{-3} at 15 dB and from 3.05×10^{-3} to 3.95×10^{-3} at 25 dB. At 30 and 35 dB, the BER increases from 3.15×10^{-3} and 3.05×10^{-3} to 3.55×10^{-3} and 3.70×10^{-3} , respectively.

These results indicate that Frobenius normalization has a non-negligible influence on the absolute BER by standardizing the effective-channel scale. Nevertheless, the overall SNR-dependent detection behavior is retained without normalization. Therefore, the normalization used in this study should be interpreted as a detector-level numerical pre-processing operation rather than as a physical received-power-preserving transformation.

Table 2. Influence of effective-channel Frobenius normalization.

Configuration	Mean Clean Power	Median Clean Power	Std. Clean Power	BER = 15 dB	BER = 25 dB	BER = 30 dB	BER = 35 dB
With Frobenius normalization	3.9943	3.2044	2.9539	7.75×10^{-3}	3.05×10^{-3}	3.15×10^{-3}	3.05×10^{-3}
Without Frobenius normalization	0.7838	0.5922	0.6719	9.20×10^{-3}	3.95×10^{-3}	3.55×10^{-3}	3.70×10^{-3}

(4) Controlled Decomposition of Channel-Aware Selection and Attention Modules

To distinguish the contribution of channel-aware spatial dimming from that of the learned attention mechanisms, a controlled decomposition was performed using four configurations: random LED selection with DeepSIC, channel-aware LED selection with DeepSIC, channel-aware selection with IIA-DeepSIC, and channel-aware selection with the complete SDDA-DeepSIC framework.

In summary, As shown in Figure 2 the controlled decomposition demonstrates that the performance improvement originates from two distinguishable mechanisms. Channel-aware LED selection substantially improves the spatial separability of the effective MIMO channel and suppresses the pronounced error floor observed with random LED activation, while IIA provides an additional gain through interference-aware user modeling. The results also reveal that the attention modules are not simply additive, since the complete SDDA configuration does not consistently outperform IIA alone. Therefore, channel-aware spatial-dimming control and IIA constitute the two most clearly supported contributors to the observed BER improvement, whereas the coupling between spatial and interference attention remains an important direction for further architectural refinement.

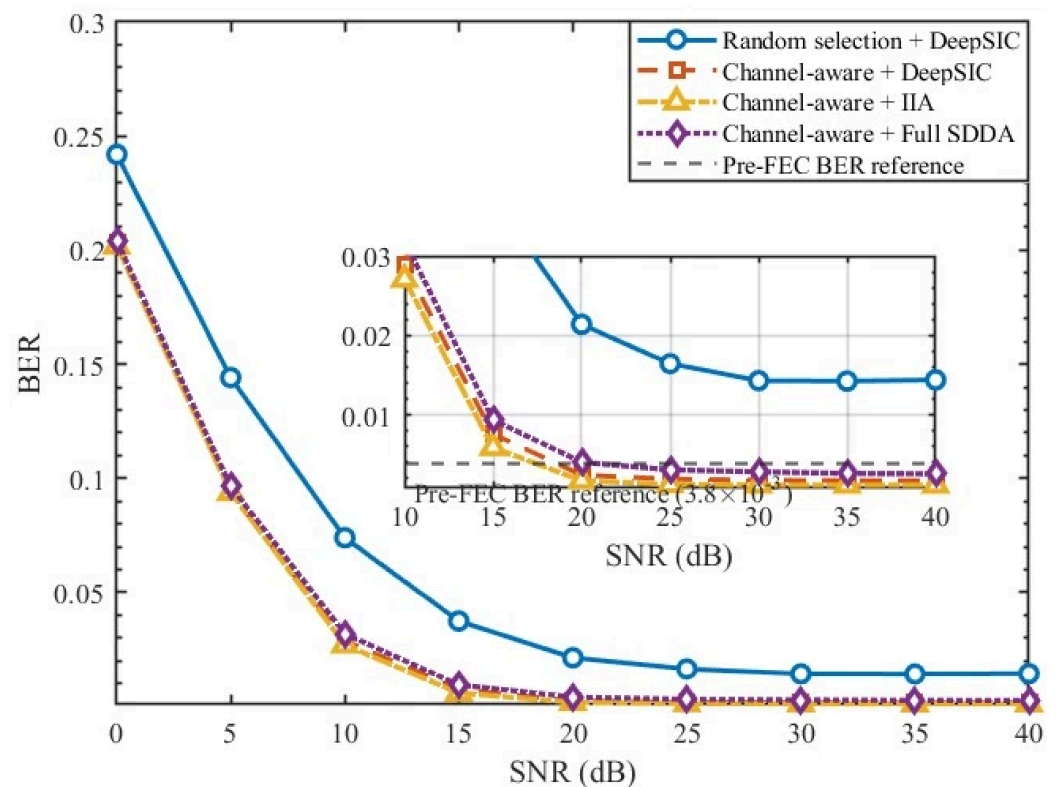


Figure 2. BER performance of the controlled decomposition under different LED-selection and attention configurations.

4.2.2. Convergence Analysis of Training

As shown in Figure 3 The initial simulation results demonstrate the sound theoretical effectiveness and learning capability of the proposed SDDA-DeepSIC framework. Both the training and validation losses exhibit stable convergence, indicating that the model can effectively learn the nonlinear detection mapping in multi-user MIMO-VLC systems. The BER results further show that the detection performance of SDDA-DeepSIC improves consistently with increasing SNR and exhibits advantages over DeepSIC and MMSE in the moderate-to-high SNR regime. These results confirm the positive contributions of spatial-dimming awareness and inter-user interference modeling to multi-user detection. In addition, SDA learns nonuniform spatial attention weights across different LEDs, while IIA captures differentiated interference-attention patterns for different target users, further supporting the effectiveness and interpretability of the dual-attention mechanism at the feature-representation level.

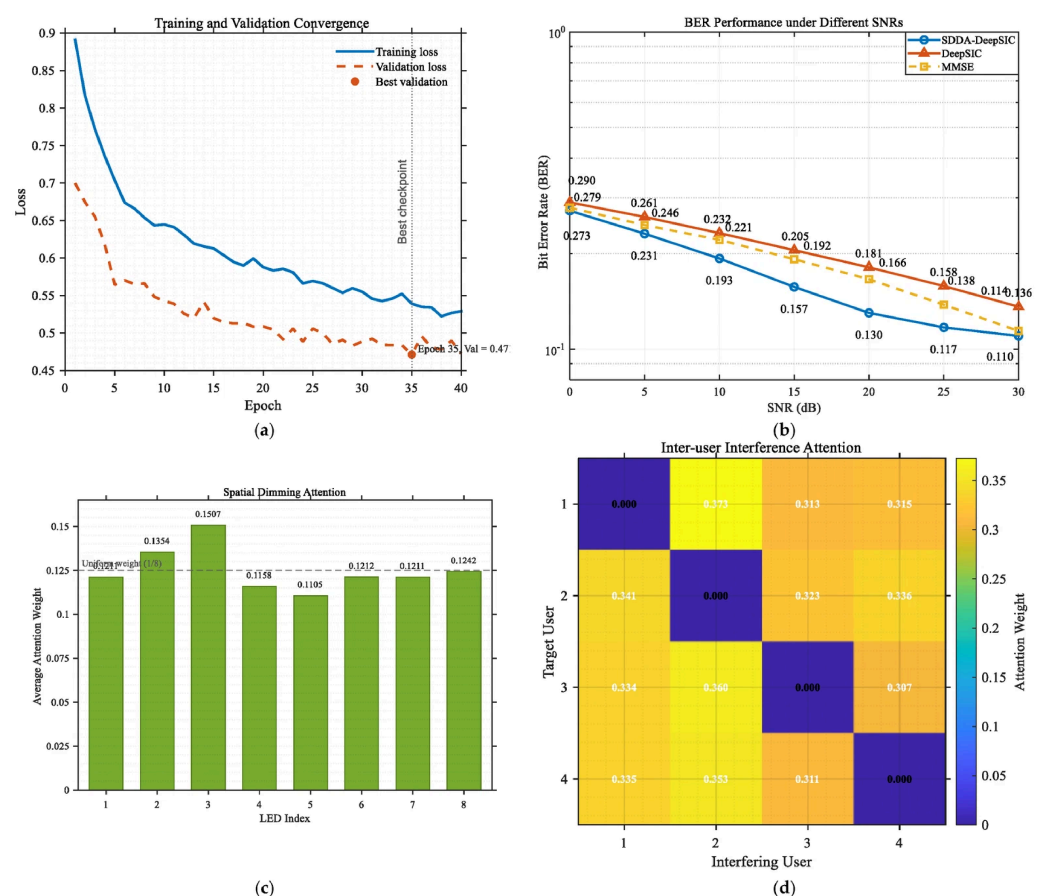


Figure 3. Analysis of the effectiveness of theoretical models. (a) Training set and validation set loss convergence curve graph. (b) Comparison of BER results between SDDA DeepSIC and MMSE. (c) Spatial dimming attention weights for different LEDs. (d) User interference attention heatmap.

Building upon these encouraging theoretical results, the channel and spatial dimming models are further refined to enhance the physical fidelity of the simulation and to evaluate the proposed framework under operating conditions that are more relevant to practical VLC systems. Specifically, a physically based Lambertian LOS channel model is introduced, the spatial dimming ratio is strictly maintained at 50%, and a channel-aware LED selection strategy based on the minimum singular value is adopted to improve the spatial separability of the effective MIMO channel. These refinements preserve the fundamental design principles of the original SDDA-DeepSIC framework while extending its performance

evaluation from theoretical validation toward a practically relevant pre-FEC operating regime, with BER values approaching or falling below the commonly adopted reference of 3.8×10^{-3} .

4.2.3. Convergence Analysis of Training After Correction

Firstly, research and analysis were conducted on the convergence of the model, and the results are shown in Figure 4. Figure 4 illustrates the training and validation convergence of the proposed model under the BER-oriented optimization objective. Both losses decrease rapidly during the initial training stage, indicating that the network efficiently learns the dominant detection and interference-related representations.

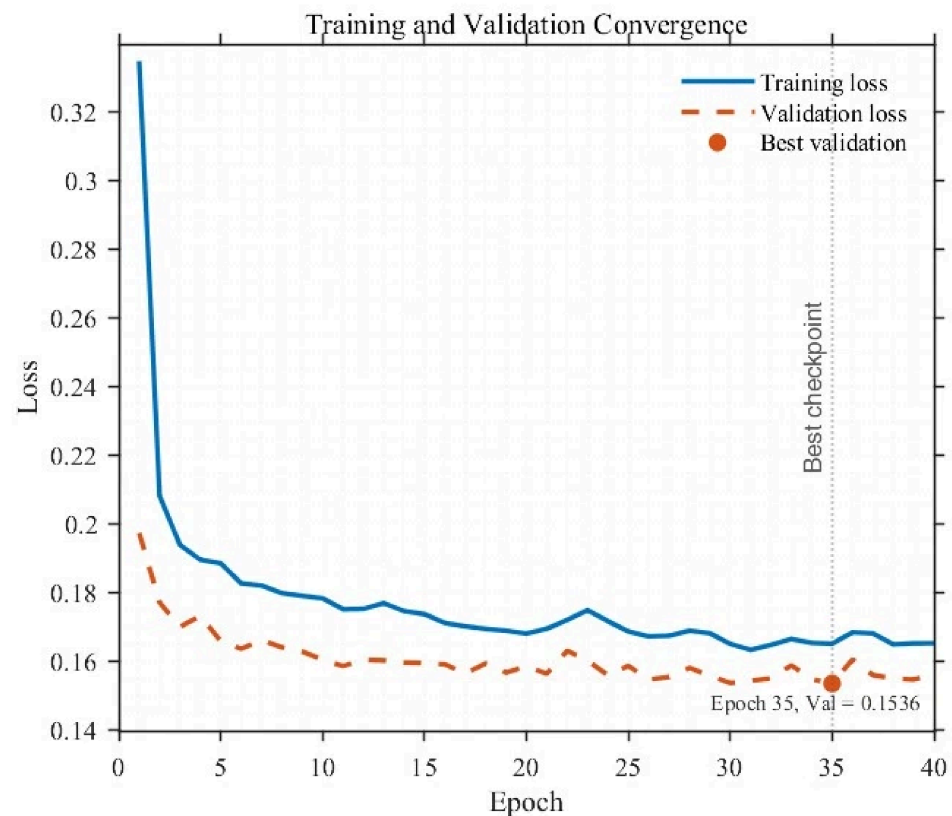


Figure 4. Training set and validation set loss convergence curve graph.

After approximately 10 epochs, the optimization enters a relatively stable fine-tuning stage, with only moderate fluctuations observed in the validation loss. The minimum validation loss of 0.15357 is obtained at epoch 35, and the corresponding model parameters are therefore selected as the final checkpoint. No pronounced divergence between the training and validation curves is observed in the later training stage, demonstrating stable convergence of the optimization procedure.

4.2.4. BER Performance Analysis

Figure 5 compares the BER performance of SDDA-DeepSIC, the original DeepSIC detector, and the conventional MMSE detector under channel-aware spatial dimming. In general, the BER of all three detectors decreases as the SNR increases, while distinct performance characteristics can be observed across different SNR regimes. At low SNRs, the performance differences among the three detectors are relatively limited because the detection process is predominantly affected by additive noise. As the SNR increases to the moderate range, SDDA-DeepSIC exhibits improved detection performance. At 10 dB, SDDA-DeepSIC achieves a BER of 3.05×10^{-2} , compared with 3.14×10^{-2} for

DeepSIC and 4.39×10^{-2} for MMSE. At 15 dB, the BER is further reduced to 7.80×10^{-3} , corresponding to reductions of approximately 6.59% and 46.76% relative to DeepSIC and MMSE, respectively. These results indicate that the attention-enhanced detector can provide useful interference-aware feature representations in the moderate-SNR regime.

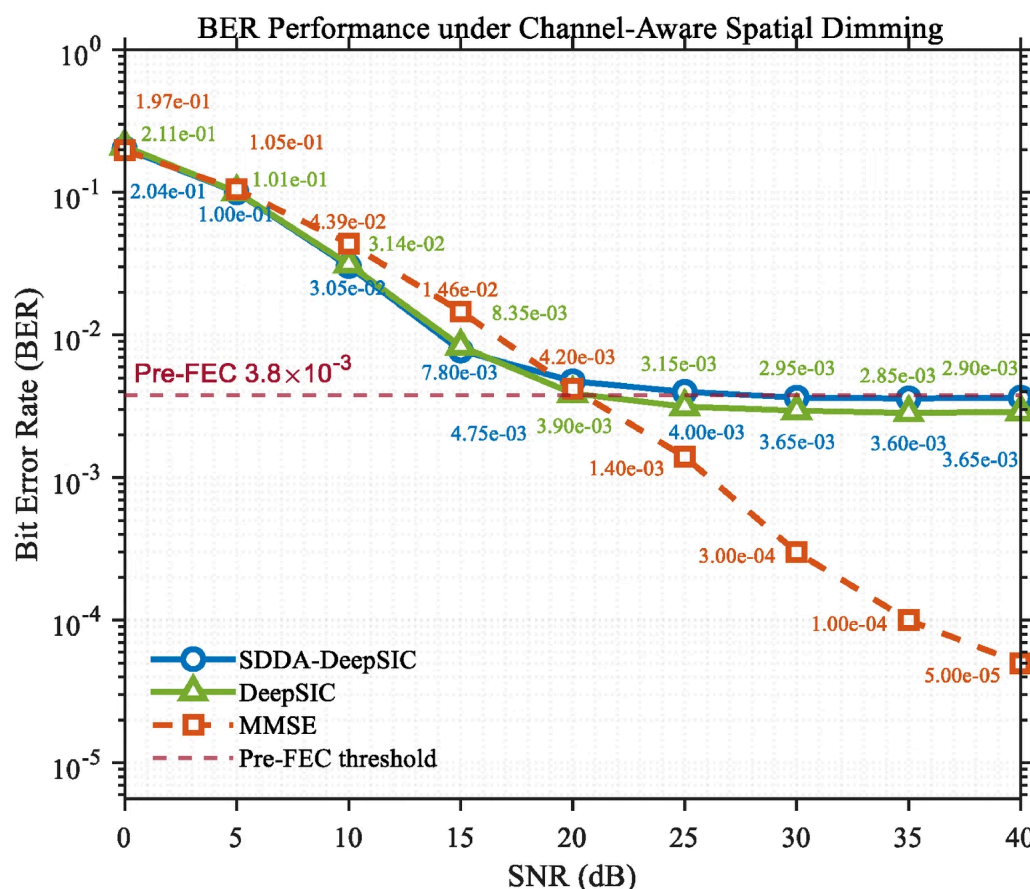


Figure 5. Comparison of BER results between SDDA DeepSIC and MMSE.

However, the performance trend changes at higher SNRs. Beyond approximately 20 dB, SDDA-DeepSIC exhibits a residual BER floor of approximately 3.6×10^{-3} – 4.0×10^{-3} , whereas DeepSIC converges to a lower BER level of approximately 2.9×10^{-3} . In contrast, the BER of MMSE continues to decrease with increasing SNR, reaching 5.0×10^{-5} at 40 dB. This behavior suggests that the learning-based detectors become increasingly limited by residual decision errors and model-dependent interference processing rather than additive noise in the high-SNR regime. The measured ablation results further show that the IIA module provides substantial BER reduction, whereas the current SDA-based spatial reweighting does not consistently improve high-SNR detection, motivating further refinement of the spatial-attention mechanism.

With respect to the commonly adopted pre-FEC BER reference of 3.8×10^{-3} , SDDA-DeepSIC reaches the reference level at 30 dB under the considered simulation configuration, while DeepSIC and MMSE reach it at 25 dB. Therefore, the present results demonstrate that the main advantage of SDDA-DeepSIC lies in improved learned detection in the low-to-moderate SNR regime rather than uniformly superior BER performance over the entire SNR range.

Figure 6 visualizes the average attention weights learned by the SDA module across the eight LEDs. The uniform attention reference is 0.125, whereas the learned weights exhibit a clear nonuniform distribution, ranging from 0.0887 to 0.1420. In particular, LED 5 receives the highest attention weight of 0.1420, approximately 13.6% above the

uniform reference, while LED 7 receives the lowest weight of 0.0887, approximately 29.0% below the uniform reference. These results demonstrate that the SDA module is capable of differentiating the relative importance of spatially distributed LED features rather than assigning identical weights to all transmitters. However, the measured ablation results indicate that such nonuniform spatial reweighting does not necessarily translate into a lower BER under all operating conditions. This suggests that excessive spatial feature reweighting may interfere with the spatial separability already established by channel-aware LED scheduling, motivating a more constrained or residual spatial-attention strategy.

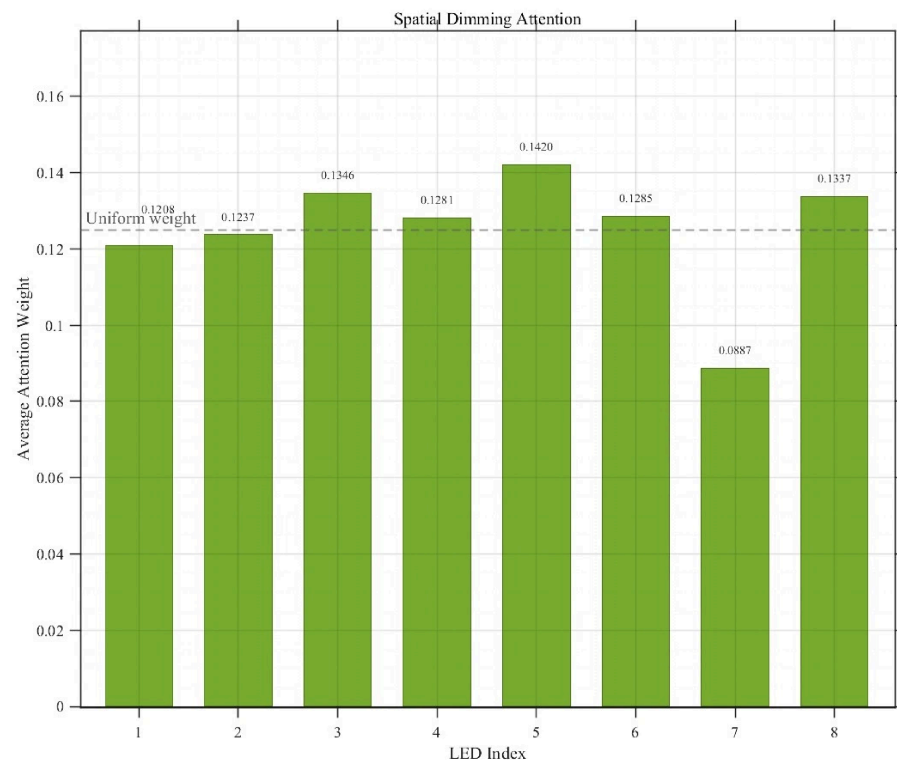


Figure 6. Spatial dimming attention weights for different LEDs.

Figure 7 visualizes the average inter-user interference attention matrix learned by the IIA module. The diagonal elements are explicitly set to zero to exclude self-interference, while the attention coefficients associated with the remaining interfering users are normalized to approximately one for each target user. In contrast to uniform interference weighting, i.e., $1/(K - 1) = 0.333$ for the considered four-user system, the learned coefficients exhibit clear user-dependent variations. For example, for target User 2, User 1 receives the largest attention coefficient of 0.374, whereas User 3 receives a smaller coefficient of 0.305. Similarly, for target User 4, the highest attention is assigned to User 3 with a coefficient of 0.368. These observations indicate that the proposed IIA module adaptively distinguishes the relative importance of different interfering users instead of treating all interference sources equally.

The learned attention matrix is also not strictly symmetric, indicating that the interference importance is target-dependent. This property is consistent with the query-key-based attention formulation, in which the channel-correlation prior provides physical guidance, while the learned feature interactions determine the final user-specific attention distribution. More importantly, the interpretability results are consistent with the measured ablation study: introducing IIA alone reduces the BER from 8.10×10^{-3} to 7.30×10^{-3} at 15 dB and from 3.15×10^{-3} to 2.05×10^{-3} at 25 dB, corresponding to relative reductions

of 9.88% and 34.92%, respectively. These results demonstrate that explicitly modeling heterogeneous inter-user interference provides effective information for DeepSIC-based multi-user detection.

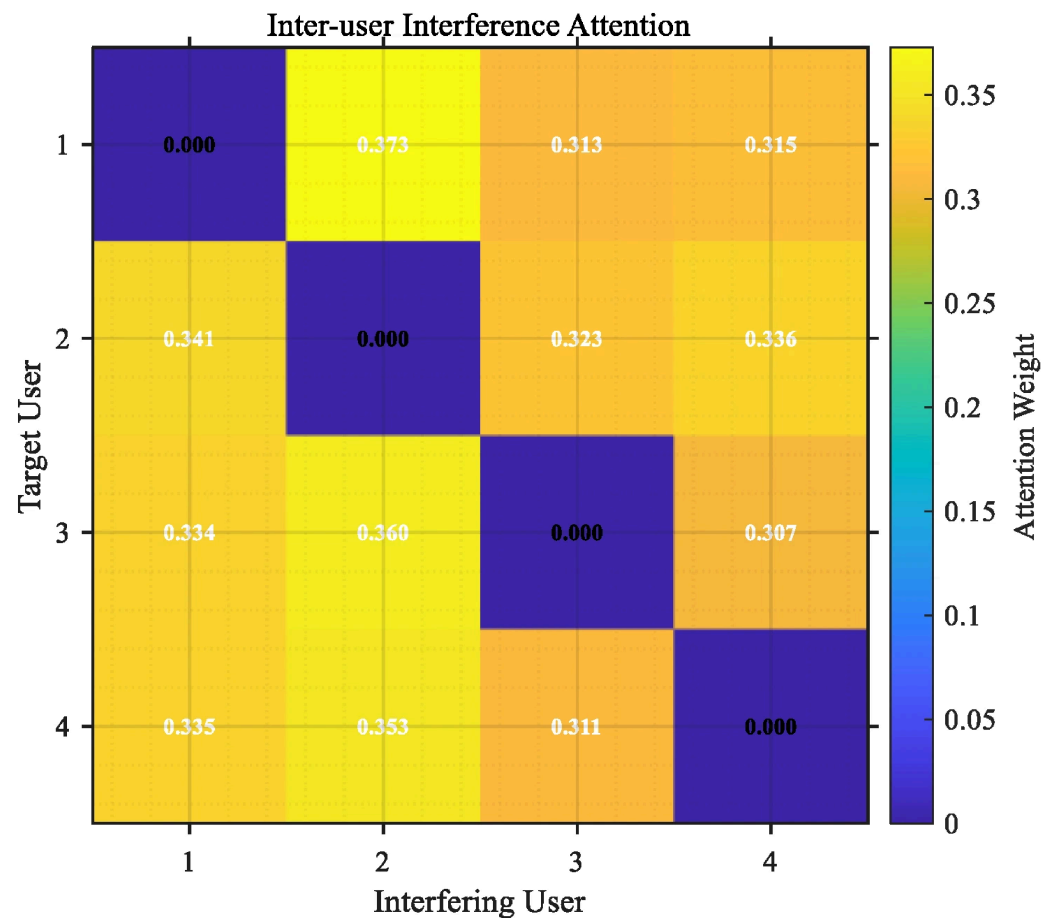


Figure 7. User interference attention heatmap.

Figure 8 illustrates the average fusion coefficients learned by the AFG across different feature dimensions. The coefficients exhibit a broad distribution, approximately ranging from 0.04 to 0.81, rather than concentrating around the equal-fusion reference of 0.5. According to $F_{fusion} = \gamma_k \odot F_s + (1 - \gamma_k) \odot F_{1,k}$, larger gate coefficients emphasize spatial-dimming features, whereas smaller coefficients assign greater importance to inter-user interference features. The average gate coefficient is approximately 0.354, indicating an overall tendency toward the IIA branch while retaining substantial dimension-dependent variations. In particular, several feature dimensions exhibit coefficients above 0.7, whereas others are below 0.1, demonstrating that AFG performs feature-specific adaptive fusion rather than using a fixed weighting strategy.

This behavior is partially consistent with the measured ablation results. At 15 dB, introducing AFG reduces the BER from 9.35×10^{-3} for the SDA + IIA configuration to 8.05×10^{-3} for the full model, corresponding to a relative reduction of approximately 13.9%. However, the same advantage is not observed at 25 dB, where the BER increases from 3.70×10^{-3} to 4.25×10^{-3} . Therefore, the learned gate distribution confirms the adaptive feature-fusion capability of AFG, while the measured BER results indicate that its detection benefit remains SNR-dependent.

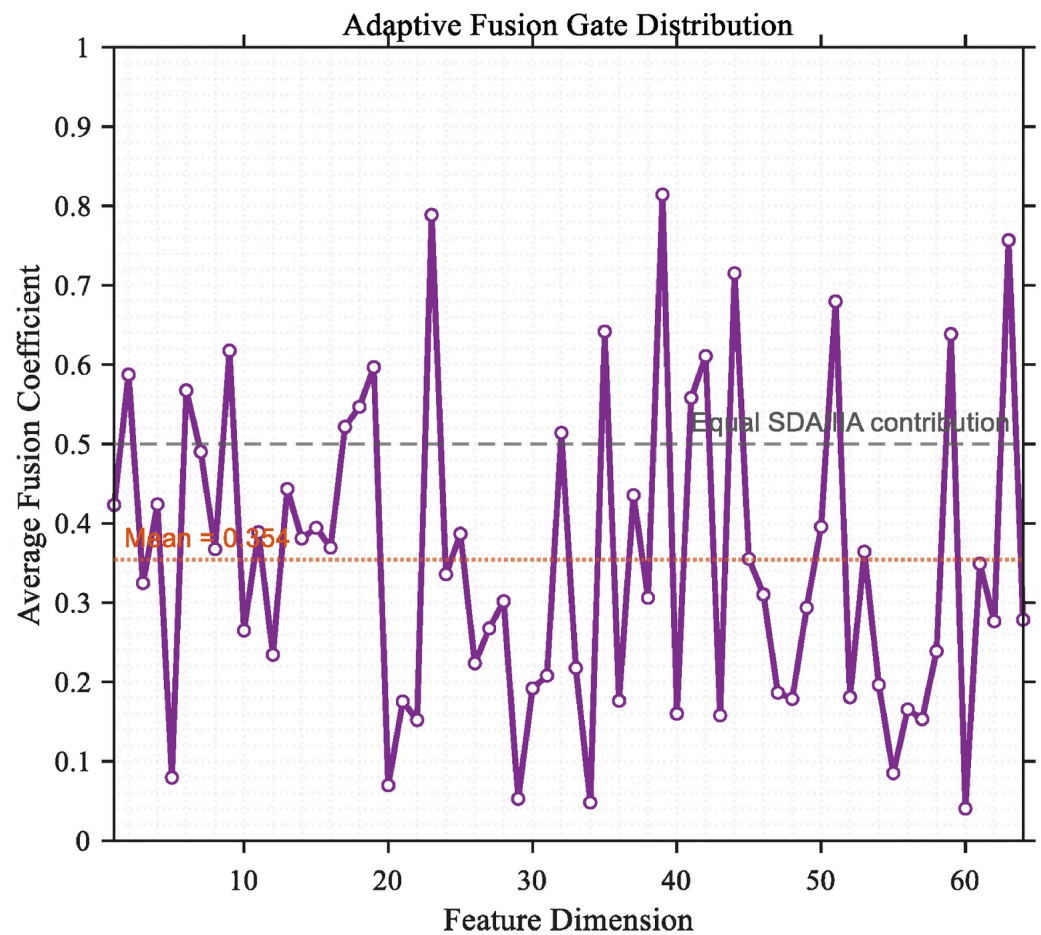


Figure 8. The distribution of adaptive fusion gates in different feature dimensions.

4.3. Ablation Study

To further verify the contribution of each key module in the proposed SDDA DeepSIC to multi-user signal detection performance, this paper designed module level ablation experiments for spatial dimming attention (SDA), inter-user interference attention (IIA), and adaptive fusion gate (AFG). The experiment uses the original DeepSIC as the benchmark model, gradually introducing different functional modules while maintaining consistency in training data, test data, network deployment layers, optimizer, learning rate, and other training hyperparameters to ensure fairness in experimental comparison. Specifically, five model configurations are set up:

1. DeepSIC: Basic model, excluding SDA, IIA, and AFG, used to provide performance benchmarks;
2. DeepSIC-SDA: Based on DeepSIC, only SDA is introduced, and LED level attention representation is established through LED activation status, spatial dimming, and related optical features to verify the effectiveness of the spatial dimming perception mechanism;
3. DeepSIC-IIA: Introducing only IIA to model the attention of potential interfering users corresponding to different target users, in order to verify the contribution of modeling non-uniform interference relationships between users;
4. SDDA w/o AFG: Simultaneously introducing SDA and IIA, but removing AFG, using direct fusion to combine the two types of attention features, for analyzing the complementary effects of dual attention mechanisms and the necessity of AFG;

5. SDDA DeepSIC: Using complete SDA, IIA, and AFG, dynamic fusion of spatial dimming features and user interference features is achieved through adaptive gating, which is the complete model proposed in this paper.

Figure 9 presents the measured module-level ablation results under identical training and testing conditions. All variants exhibit a rapid BER reduction as the SNR increases, whereas their performance differences become more pronounced in the moderate-to-high SNR regime. Among the investigated modules, IIA provides the most significant and consistent improvement. At 15 dB, incorporating IIA into the original DeepSIC reduces the BER from 8.10×10^{-3} to 7.30×10^{-3} , corresponding to a relative reduction of 9.88%. At 25 dB, the BER is further reduced from 3.15×10^{-3} to 2.05×10^{-3} , yielding a 34.92% improvement. The advantage of the IIA-enhanced detector remains evident in the high-SNR regime, demonstrating the effectiveness of explicitly modeling heterogeneous inter-user interference.

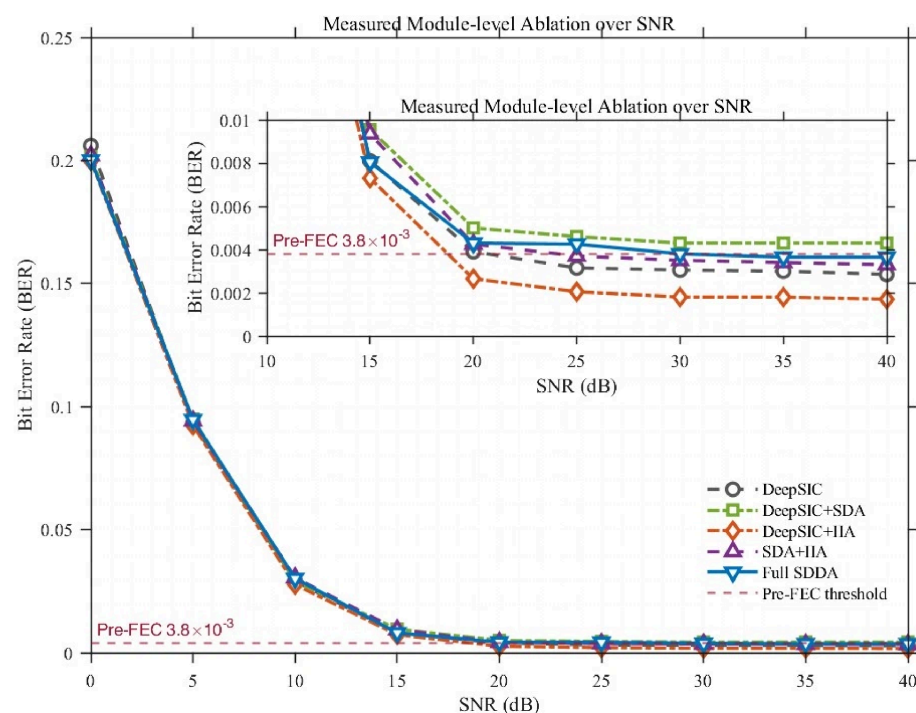


Figure 9. Module-level ablation over SNR.

In contrast, introducing SDA alone increases the BER to 9.55×10^{-3} at 15 dB and 4.60×10^{-3} at 25 dB, corresponding to degradations of 17.90% and 46.03%, respectively, relative to the original DeepSIC. Although the learned SDA weights exhibit nonuniform LED-dependent spatial attention, the ablation results indicate that direct spatial reweighting does not necessarily translate into improved detection performance. One possible explanation is that additional spatial reweighting may partially perturb the spatial separability already established by the channel-aware LED scheduling strategy.

Moreover, directly combining SDA and IIA does not outperform the IIA-only configuration, indicating that the two attention branches are not simply additive and that interactions between spatial and interference-aware representations must be carefully controlled. At 15 dB, introducing AFG reduces the BER of the SDA + IIA configuration from 9.35×10^{-3} to 8.05×10^{-3} , corresponding to a relative improvement of approximately 13.9%. However, this benefit is not maintained at 25 dB, where the BER increases from 3.70×10^{-3} to 4.25×10^{-3} . These observations suggest that AFG can mitigate unfavor-

able feature interactions in the moderate-SNR regime, whereas its effectiveness remains SNR-dependent.

Overall, the measured ablation results identify inter-user interference modeling as the principal source of performance improvement in the current architecture. They also reveal that unconstrained spatial reweighting and subsequent dual-branch fusion may limit the high-SNR performance, providing a clear motivation for developing a residual or constrained spatial-attention mechanism.

Figures 10 and 11 further compare the measured BER of different module configurations at representative SNRs of 15 and 25 dB. At 15 dB, the original DeepSIC achieves a BER of 8.10×10^{-3} . Incorporating SDA alone increases the BER to 9.55×10^{-3} , whereas introducing IIA reduces it to 7.30×10^{-3} , corresponding to a relative BER reduction of 9.88%. Directly combining SDA and IIA results in a BER of 9.35×10^{-3} . After introducing AFG, the BER decreases to 8.05×10^{-3} , representing a 13.90% reduction relative to the ungated SDA + IIA configuration. This result indicates that AFG can partially mitigate unfavorable interactions between the two attention branches under moderate-SNR conditions.

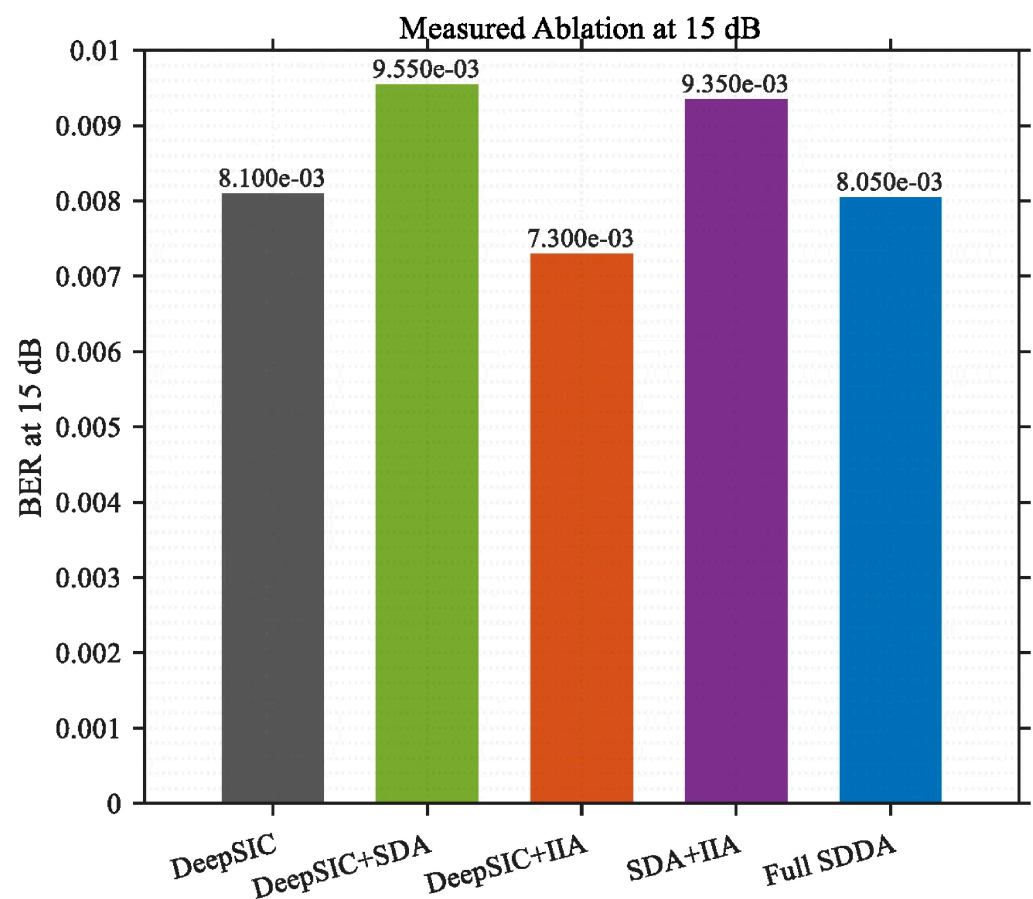


Figure 10. Measured ablation at 15 dB.

At 25 dB, the differences among the investigated modules become substantially more pronounced. IIA achieves the lowest BER of 2.05×10^{-3} , corresponding to a 34.92% reduction relative to the 3.15×10^{-3} obtained by the original DeepSIC. In contrast, SDA increases the BER to 4.60×10^{-3} , while the SDA + IIA and full SDDA configurations yield 3.70×10^{-3} and 4.25×10^{-3} , respectively. With respect to the pre-FEC BER reference of 3.8×10^{-3} , the IIA-enhanced detector exhibits the largest BER margin at 25 dB. These results demonstrate that explicit inter-user interference modeling constitutes the dominant source of performance improvement in the current architecture, particularly as the system transitions from a noise-dominated to an interference/model-limited operating regime.

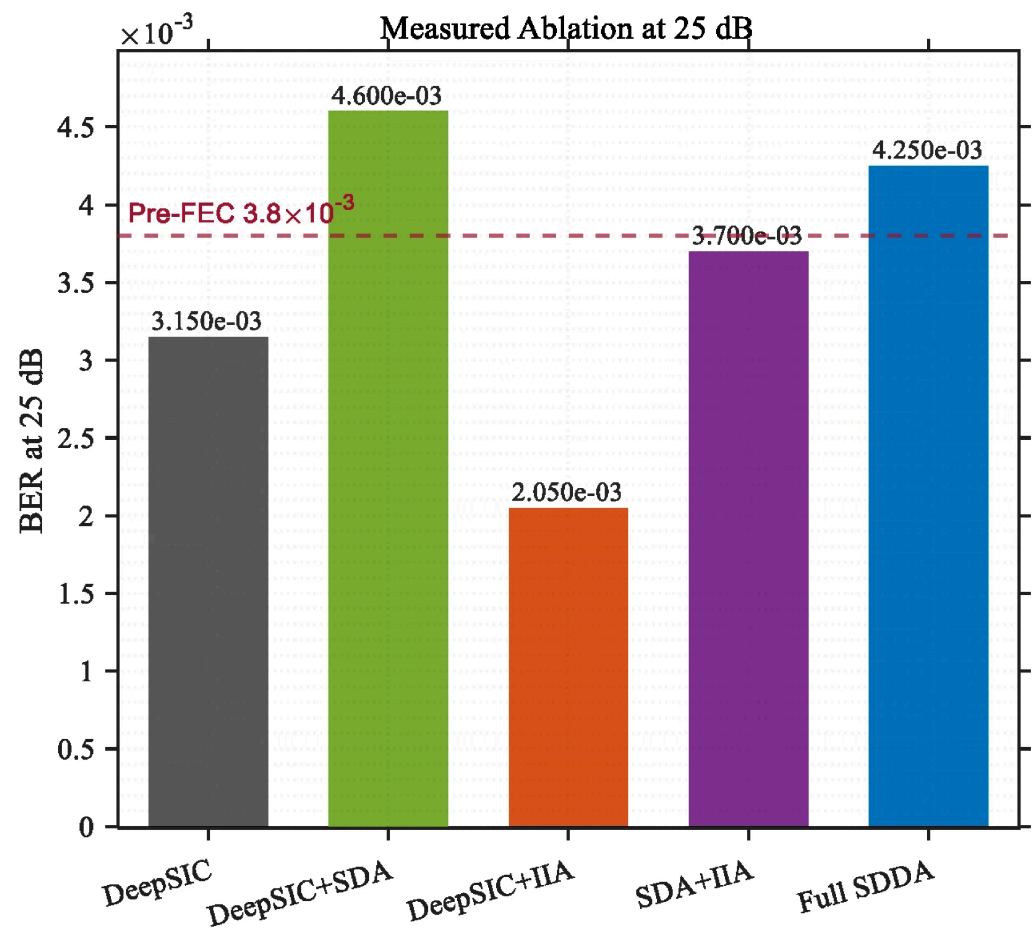


Figure 11. Measured ablation at 25 dB.

As shown in Table 3. Among them, "-" represents modules that do not exist, and "✓" represents modules that exist. The comparison between the two SNR points also reveals an SNR-dependent interaction among the proposed modules. The relative BER reduction provided by IIA increases from 9.88% at 15 dB to 34.92% at 25 dB, whereas the degradation associated with SDA increases from 17.90% to 46.03%. Moreover, although AFG improves the SDA + IIA configuration at 15 dB, the same benefit is not observed at 25 dB. This observation suggests that direct spatial reweighting and adaptive dual-branch fusion require further constraints to preserve the spatial separability established by channel-aware LED scheduling in the high-SNR regime.

Table 3. Measured ablation results of different module configurations.

Model	SDA	IIA	AFG	Best Epoch	Best Val. Loss	BER @ 15 dB	BER Gain @ 15 dB (%)	BER @ 25 dB	BER Gain @ 25 dB (%)	Params	Latency (ms)
DeepSIC	—	—	—	28	0.15354	0.00810	0	0.00315	0	25,605	22.427
DeepSIC + SDA	✓	—	—	24	0.15641	0.00955	−17.90	0.00460	−46.03	30,886	42.788
DeepSIC + IIA	—	✓	—	27	0.15494	0.00730	+9.88	0.00205	34.92	30,917	36.646
SDA + IIA	✓	✓	—	24	0.15606	0.00935	−15.43	0.00370	−17.46	32,038	58.654
Full SDDA	✓	✓	✓	35	0.15357	0.00805	+0.62	0.00425	−34.92	40,294	79.177

Overall, the measured ablation study provides several important observations regarding the roles of the proposed modules. First, IIA is identified as the dominant contributor to the detection improvement. Compared with the original DeepSIC, IIA reduces the BER by 9.88% at 15 dB and 34.92% at 25 dB, demonstrating that explicit

modeling of heterogeneous inter-user interference becomes increasingly beneficial as the influence of additive noise decreases. Second, the results reveal that spatial-dimming awareness and interference-aware modeling are not simply additive. Although SDA learns nonuniform LED-dependent attention patterns, direct spatial reweighting does not consistently improve BER under channel-aware LED scheduling, suggesting a non-trivial coupling between spatial-dimming features and the effective channel structure. Third, AFG partially alleviates this feature interaction in the moderate-SNR regime, reducing the BER of the SDA + IIA configuration by 13.90% at 15 dB. These observations indicate that interference-aware modeling provides the most robust performance gain, while spatial-dimming feature refinement and adaptive fusion exhibit SNR-dependent behavior and offer opportunities for further architectural improvement.

Among all evaluated variants, DeepSIC + IIA achieves the lowest BER at both representative SNR points, reducing the BER by 9.88% at 15 dB and 34.92% at 25 dB relative to the original DeepSIC, while introducing only a moderate increase in model complexity.

To investigate the origin of the high-SNR BER floor, we further analyze the layer-wise BER, normalized residual multi-user interference, and the stability of the learned attention coefficients. The results in Figure 12 show that the residual interference of the complete SDDA model gradually approaches a non-zero plateau at high SNR. For example, at 40 dB, the normalized residual interference after the fifth unfolding layer is 4.97×10^{-3} , compared with 3.37×10^{-3} for IIA-DeepSIC. This behavior is consistent with the corresponding BERs of 3.80×10^{-3} and 2.65×10^{-3} , respectively, suggesting that residual multi-user interference becomes increasingly important once the additive-noise contribution is sufficiently reduced.

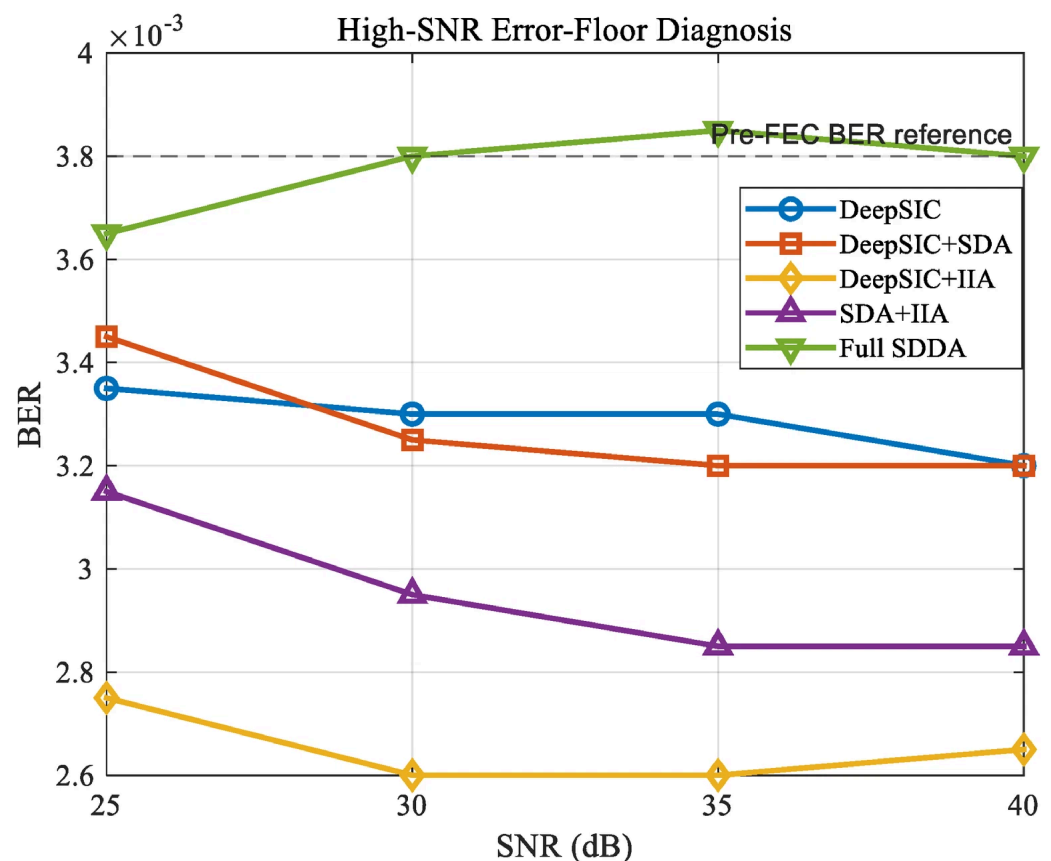


Figure 12. High-SNR error-floor analysis.

In contrast, the learned attention coefficients remain highly stable over the investigated high-SNR range. The IIA coefficient of variation changes only from 0.352 to 0.349, while the

AFG coefficient of variation remains approximately 0.596 over 25–40 dB. The mean AFG coefficient is also nearly invariant at approximately 0.354. Therefore, the observed BER saturation cannot be attributed to an SNR-dependent instability of the attention weights. Combined with the module-level ablation results, these observations indicate that the high-SNR error floor is more closely associated with persistent residual interference and the interaction between spatial reweighting and interference-aware feature representations.

5. Conclusions

This paper developed a spatial-dimming- and interference-aware DeepSIC framework for multi-user MIMO-VLC systems. A physically based Lambertian channel model and channel-aware LED selection were employed to preserve effective channel separability under spatial-dimming constraints. SDA, IIA, and AFG were further introduced to model spatial-dimming features, heterogeneous inter-user interference, and their adaptive interactions, respectively. Numerical simulations demonstrate effective multi-user detection in the low-to-moderate SNR regime. More importantly, the measured ablation study identifies IIA as the primary source of performance improvement, reducing the BER by 9.88% and 34.92% at 15 and 25 dB, respectively, relative to the original DeepSIC. The results further reveal that spatial-dimming- and interference-aware representations are not simply additive, while adaptive fusion exhibits SNR-dependent behavior. Overall, the results highlight the strong coupling among spatial dimming, effective channel structure, and inter-user interference, and demonstrate the importance of explicit interference-aware modeling for learned SIC detection. Future work will investigate physically constrained spatial attention, SNR-aware adaptive fusion, NLOS propagation, imperfect CSI, and hardware-based VLC validation. The present evaluation is based on a controlled LOS-dominant Lambertian channel model. Higher-order reflections, multipath propagation, dynamic blockage, receiver orientation/FOV variations, and optical front-end nonlinearities may modify the effective channel and, consequently, the interference topology learned by IIA. Future work will therefore investigate the proposed detector under NLOS/multipath channels, dynamic blockage, imperfect CSI, receiver orientation variations, and nonlinear LED characteristics, together with hardware-based VLC measurements. The present performance evaluation is limited to BPSK-based transmission. Extension to higher-order PAM and optical OFDM, together with modulation-dependent nonlinear and optical front-end effects, constitutes an important direction for future work.

Author Contributions: Conceptualization, J.W.; methodology, J.W. validation, J.W.; data curation, J.W.; writing—original draft preparation, J.W.; writing—review and editing, G.C.; visualization, G.C.; project administration, G.C.; funding acquisition, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: Technical foundation (JZX7J202602ZL012100), Changchun Science and Technology Bureau (25GNYZ35).

Institutional Review Board Statement: This study did not involve humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank all of the authors cited and the anonymous reviewers in this article for their helpful suggestions.

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

References

- Altunbas, Y.; Turk, K. Sum-rate maximization for downlink rate-splitting multiple access (RSMA) in visible light communication (VLC). *Opt. Quantum Electron.* **2026**, *58*, 170. [CrossRef]
- Tugcu, E. Mitigation red-shift effect in multi-color LEDs-based visible light communication with signal-dependent noise. *Phys. Commun.* **2025**, *69*, 102601. [CrossRef]
- Al, I.; Chowdhury, M.Z.; Jang, Y.M. Pyramid Attention Residual UNet-Based Channel Estimation of Indoor Massive MIMO Visible Light Communication System. *IEEE Access* **2025**, *13*, 198677–198687. [CrossRef]
- Mehrpooya, A.; Yokar, V.N.; Teng, Y.; Shen, S.; Zhao, W.; Olianeshad, A.; Bahrani, S.; Emami, A.; Zhang, N.; Yan, S.; et al. Delay-Aware Time-Coordinated Nanosecond Optical Switching for Dynamically Reconfigurable Optical Networks. *Proceedings of the 2026 International Conference on Optical Network Design and Modelling (ONDM)* **2026**.
- Givary, J.; Almusawwir, M.N.; Nabila, A.P.; Anshari, M.D.; Nugroho, A.; Hambali, A. Analisis Daya Terima Pada Visible Light Communication Berbasis Multiple-Input dan Multiple-Output. 2026. Available online: https://www.researchgate.net/publication/299452179_Analisis_daya_terima_pada_Visible_Light_Communication_berbasis_Multiple-Input_dan_Multiple-Output (accessed on 8 October 2026).
- Lu, H.H.; Lin, Y.P.; Wu, P.Y.; Chen, C.Y.; Chen, M.C.; Jhang, T.W. A multiple-input-multiple-output visible light communication system based on VCSELs and spatial light modulators. *Opt. Express* **2014**, *22*, 3468–3474. [CrossRef]
- Geto, H.G.; Tola, G.T. Enhancement in power stratification efficiency of NOMA-MIMO-aided downlink indoor visible light communication networks. *Opt. Commun.* **2025**, *579*, 131499. [CrossRef]
- Xia, J.; Zhu, B.; Zhang, Z. Decorrelation Lenses for Massive MIMO Visible Light Communications. *Photonics Technol. Lett. IEEE* **2025**, *37*, 575–578. [CrossRef]
- Su, X.; Yu, H.F.; Kim, W.; Choi, C.; Choi, D. Interference cancellation for non-orthogonal multiple access used in future wireless mobile networks. *Eurasip J. Wirel. Commun. Netw.* **2016**, *2016*, 231. [CrossRef]
- Alegra, J.V.; Sheikhi, A.; Edfors, O. Zero-Forcing MU-MIMO Precoding under Power Amplifier Non-Linearities. In *Proceedings of the 2025 59th Asilomar Conference on Signals, Systems, and Computers*; IEEE: Piscataway, NJ, USA, 2025.
- Kumar, N.; Kumar, K. Performance Analysis of M*N Equalizer based Minimum Mean Square Error (MMSE) Receiver for MIMO Wireless Channel. *Int. J. Comput. Appl.* **2011**, *6*, 47–50. [CrossRef]
- Harshavardhan, P.; Raja, A. Estimation of channel using novel linear minimum mean square error method in comparison with linear filtering method for MIMO-OFDM communication systems. *AIP Conf. Proc.* **2024**, *3097*, 020017. [CrossRef]
- Chen, Z.; Huang, S.; Jiang, M. Two-Path Serial Inter-Symbol-Interference Cancellation for Air-to-Ground Communication. *Commun. Lett. IEEE* **2025**, *29*, 100–104. [CrossRef]
- Nammi, S. Single Codeword, Multi-Layer Serial Interference Cancellation (Sic) for Spatial Multiplexing. US20200162139A1, 7 August 2026.
- Zhu, X.; Murch, R.D. Performance analysis of maximum likelihood detection in a MIMO antenna system. *IEEE Trans. Commun.* **2002**, *50*, 187–191. [CrossRef]
- Meng, F.; Liu, Y.; Wang, L.; Zhou, W.; Zhou, X. Low-depth quantum approximate optimization algorithm for maximum likelihood detection in massive MIMO. *Quantum Inf. Process.* **2025**, *24*, 294. [CrossRef]
- Akpınar, E.; Aslandogan, E.; Ozden, B.A.; İlhan, H.; Aydın, E. Deep Learning-Enabled Signal Detection for MIMO-OTFS-Based 6G and Future Wireless Networks. *arXiv* **2026**, arXiv:2601.13635.
- Kumar, A.; Sharan, N.; Dixit, V.; Gaur, J.; Kumar, R.; Dass, S.; Kumar, A. Adaptive Detection Strategies for Enhanced BER Performance in MIMO Visible Light Communication Systems. *J. Opt.* **2026**. [CrossRef]
- Feng, R.H.; Wei, Y.D.; Khan, W.; Xie, S.; Luo, L.-P.; Yousaf, M.Z. Indoor massive MIMO-VLC precoding under SDN-FSO channel. *Phys. Commun.* **2025**, *73*, 102847. [CrossRef]
- Feng, R.; Ma, X.; Qian, L.; Xie, S.; Zhang, W.; Luo, L. Tri-Level BCD Optimization in OIRS-Assisted MIMO-VLC for Enhanced Physical-Layer Security. *Commun. Lett. IEEE* **2025**, *29*, 2233–2237.
- Amin, M.; Tanveer, T.A.; Shah, S.; Shafi, M. DeepSIC: A deep model. *Clust. Comput.* **2018**, *21*, 741–754. [CrossRef]
- Luo, S.; Yang, Y.; Yin, Y.; Shen, C.; Zhao, Y.; Song, M. *DeepSIC: Deep Semantic Image Compression*; Springer: Cham, Switzerland, 2018.
- Shlezinger, N.; Fu, R.; Eldar, Y.C. DeepSIC: Deep Soft Interference Cancellation for Multiuser MIMO Detection. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 1349–1362. [CrossRef]
- Raviv, T.; Park, S.; Simeone, O.; Shlezinger, N. Modular Model-Based Bayesian Learning for Uncertainty-Aware and Reliable Deep MIMO Receivers. In *Proceedings of the 2023 IEEE International Conference on Communications Workshops (ICC Workshops)*; IEEE: Piscataway, NJ, USA, 2023.
- Wu, K.; Jiang, W.; Niyato, D.; Li, Y.; Luo, C. Transfer-based Adversarial Poisoning Attacks for Online (MIMO-)Deep Receivers. *arXiv* **2024**, arXiv:2409.02430.

26. Marshoud, H.; Muhaidat, S.; Sofotasios, P.C.; Hussain, S.; Imran, M.A.; Sharif, B.S. Optical Non-Orthogonal Multiple Access for Visible Light Communication. *IEEE Wirel. Commun.* **2018**, *25*, 82–88. [\[CrossRef\]](#)
27. Shao, S.; Khreishah, A.; Ayyash, M.; Rahaim, M.B.; Elgala, H.; Jungnickel, V.; Schulz, D.; Little, T.D.C.; Hilt, J.; Freund, R. Design and Analysis of a Visible-Light-Communication Enhanced WiFi System. *IEEE/OSA J. Opt. Commun. Netw.* **2015**, *7*, 960–973. [\[CrossRef\]](#)
28. Caregnato, N.A.; Maximo, M.R.O.A.; Afonso, R.J.M. Mixed-Integer Motion Planning for Nonholonomic Robots Under Visible Light Communication Constraints. *Optim. Control-Appl. Methods* **2025**, *46*, 1023–1035. [\[CrossRef\]](#)
29. Jung, S.H.; Lee, K.; Lee, H.; Choi, R.; Lee, J. Fast-Response and High-Color-Purity RGB Fluorescent Organic Light-Emitting Diodes for Visible Light Communication System. *Adv. Opt. Mater.* **2025**, *13*, 2402693. [\[CrossRef\]](#)
30. Yokar, V.N.; Le-Minh, H.; Li, X.; Woo, W.L.; Alves, L.N.; Zvanovec, S.; Son, T.T.; Ghassemlooy, Z. A Novel Frame Identification and Synchronization Technique for Smartphone Visible Light Communication Systems Based on Convolutional Neural Networks. *arXiv* **2025**, arXiv:2506.23004.
31. Zarini, H.; Mohammadisarab, A.; Dehkordi, M.F.; Mili, M.R.; Safaei, B.; Movaghar, A.; Coleri, S.; Jorswieck, E. On the Performance of Unmanned Aerial Vehicles with MIMO VLC. In *Proceedings of the ICC 2025-IEEE International Conference on Communications*; IEEE: Piscataway, NJ, USA, 2024.
32. Wu, M.; Yao, T.; Xing, Z.; Zhang, J.; Lin, Y. Channel Estimation Algorithm with Fuzzy-Enhanced Compressed Sensing for Indoor MIMO VLC. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2025**, *e108.a*, 986–990.
33. Mahyastuty, V.W.; Iskandar, H.; Arifianto, M.S. Adaptive Successive Interference Cancellation using Deep Learning for High Altitude Platform Station in Various K-Rician Channel. *IAENG Int. J. Comput. Sci.* **2022**, *49*, 495–500.
34. Hassan, E.S.; Jabbari, A.; Alharbi, A.A. Evaluating MIMO-VLC System Performance: Modulation Techniques and Ambient Light Interference Effects. *Photonics* **2025**, *12*, 649. [\[CrossRef\]](#)
35. Ibrahim, H.S.; Abaza, M.; Mansour, A.; Alfalou, A. Performance Analysis of Power Allocation and User-Pairing Techniques for MIMO-NOMA in VLC Systems. *Photonics* **2024**, *11*, 206. [\[CrossRef\]](#)
36. Sikder, P.; Rahman, M.T.; Bakibillah, A.S.M. Advancements and Challenges of Visible Light Communication in Intelligent Transportation Systems: A Comprehensive Review. *Photonics* **2025**, *12*, 225. [\[CrossRef\]](#)
37. Salama, W.M.; Aly, M.H.; Amer, E.S. Deep learning based BER improvement for NOMA-VLC systems with perfect and imperfect successive interference cancellation. *Opt. Quantum Electron.* **2023**, *55*, 692. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.