Received: 24 October 2025 | Revised: 27 April 2026 | Accepted: 5 August 2026 | Published online: 22 September 2026

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–12
DOI: 10.47852/bonviewJCCE62027976

RESEARCH ARTICLE

# Efficient Communication Management in Diffractive Optics IoT Networks for Intelligent Video Transmission

**Mohammed Abdullah Al-Mekhlafi[1], Mohammad Kamrul Hasan[2,*], Qusai Y. Shambour[3], Ali Q. Saeed[4], Saed Adnan Mustafa[5,6] and Taher M. Ghazal[2,7,8]**

[1] *College of Business Administration, University of Business and Technology, Saudi Arabia*

[2] *Center for Cyber Security, Universiti Kebangsaan Malaysia, Malaysia*

[3] *Department of Data Science and Artificial Intelligence, Al-Ahliyya Amman University, Jordan*

[4] *Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University, Iraq*

[5] *College of Business, The American University of Kurdistan, Iraq*

[6] *MidOcean University, Comoros*

[7] *Faculty of Computing and IT, Sohar University, Oman*

[8] *Department of Networks and Cybersecurity, Al-Ahliyya Amman University, Jordan*

**Abstract:** Dynamic, intelligent video-centric Internet of Things (IoT) networks need strong communication solutions for high data rates, low latency, and energy efficiency. Radio Frequency (RF)-based IoT frameworks are unsuitable for real-time video transmission because of interference, high retransmission costs, and restricted capacity. This paper suggested DOECOM-VID, a novel architecture that employs diffractive optical elements (DOEs) for interference-free parallel data distribution, phase-front shaping, and multi-channel optical beamforming. In crowded IoT situations, DOECOM-VID's Deep Reinforcement Learning (DRL) optimization engine and Fourier-domain diffractive Multiple-Input Multiple-Output (MIMO) channel model can identify paths, direct beams, and dynamically distribute resources. High Definition (HD) video streaming is enabled via phase-coded diffractive waveguides that improve spatial multiplexing and reduce diffraction losses. DOECOM-VID increases spectral efficiency by 31%, end-to-end latency by 34%, and Peak Signal-to-Noise Ratio (PSNR) by 40.6% over existing IoT video transmission systems. Energy-aware routing and adaptive modulation may extend IoT video node lifespan by 25%. DOECOM-VID delivers interference-resistant, high-capacity optical communication for smart city video analytics, autonomous surveillance, and Augmented Reality/Virtual Reality (AR/VR) streaming. Future expansions will investigate quantum-enhanced diffractive computing for ultra-secure real-time video transmission.

**Keywords:** diffractive optical elements (DOEs), deep reinforcement learning (DRL), Internet of Things (IoT) networks, intelligent video transmission, Fourier-domain diffractive MIMO

## 1. Introduction

Smart cities, autonomous driving, security monitoring, augmented/virtual reality, and remote healthcare are stressing the Internet of Things (IoT) infrastructure like never before. These applications require interference-resistant connectivity to send high-resolution, real-time video. In addition, they require fast data and minimal latency. While radio frequency (RF) IoT frameworks are extensively used, they struggle to enable video-centric applications in chaotic and unpredictable settings. Spectrum, signal attenuation, multipath fading, and severe interference make RF-based video transmission quality poor. These approaches are inadequate for many next-generation applications because retransmission wastes power and delays. Additionally, Budati et al. [1] improved 5G sensor network video streaming efficiency and throughput. Interference control and spectrum congestion

*Corresponding author: Mohammad Kamrul Hasan, Center for Cyber Security, Universiti Kebangsaan Malaysia, Malaysia. Email: mkhasan@ukm.edu.my

are still issues for radio frequency (RF), needing a significant communication technology revolution.

Recently, diffractive optics has been an alternative. Modifying optical wavefront phase, amplitude, and direction changes light propagation. These components are ideal for scalable, parallel, high-capacity Internet of Things connections because of their compact size, low weight, and intricate beam patterns. As previously stated, Li et al. [2] say opaque obstructions do not impede diffractive optical communication. Despite obstacles, optical communication is durable and not line-of-sight-dependent. The results allow IoT optical communication frameworks that circumvent RF system limitations.

Diffraction and refractive elements in optical systems may improve broadband imaging and data transfer. To boost broadband imaging depth of field, MiriRostami et al. [3] designed a hybrid diffractive-refractive optical system. This setup illustrates how diffractive optics handles high-quality data. Jiang et al. [4] and Li et al. [5] advocated diffusers for advanced all-optical transmission and classification systems. DOE-based systems may offer intelligent optical sensing, decoding, and secure data transport, according to these models. Even with advancements, most optical systems are for sterile laboratories or decoding/classification. Practical end-to-end IoT video network communication frameworks are understudied. These frameworks should include intelligent route optimization, diffractive beamforming, and resource management. Contemporary IoT setups need more than optical systems' adaptive intelligence. Internet of Things traffic, node mobility, and energy utilization may alter. Diffraction may cause data loss when optical beams travel through complicated materials or long distances. To help, scientists built diffractive channel and beam optimization models. They cannot adjust to network changes without integration. Model-driven and zero-power optical convolutional neural networks exhibit low IoT video transmission resolution, according to Liu et al. [6] and Fei et al. [7]. Visual IoT creates privacy and security problems. Although multi-key encoding and block scrambling are lightweight security upgrades, Hosny et al. [8] discovered that the transmission architecture is often disregarded. A reliable visual communication system involves adaptive control and enduring physical layers. The basic wave propagation constraints of diffraction, dispersion, and interference increase optical system challenges under changeable weather.

Jow et al.'s [9] comprehensive investigation of astrophysical lensing's refractive and diffractive regimes emphasized the Fresnel scale and optical domain transition. Diffractive beamforming in IoT communication systems demands coherence across several transmission channels, making these approaches important. Developing reliable optical communication needs exact propagation models, which only our study can provide. Wang et al. [10] developed Fourier optical speckle-based compressive sensing for meaningful visual transmission. Resource-constrained WSNs need good image and video encoding. Their solution lowered WSN image reconstruction bandwidth and transmission overhead without impacting video quality. Unlike this approach, they concentrate on compression rather than adaptive beam and resource management in real-time video distribution.

DRL is a physics-coupled, cross-layer optimization technique in diffractive optics, not RF, in the proposed DOECOM-VID architecture. The DRL agent uses an augmented Markov Decision Process (MDP) with network-level descriptors (queue dynamics, traffic intensity, and latency states) and optical field representations like phase-front distributions, diffraction efficiency coefficients, and Fresnel–Kirchhoff spatial beam propagation profiles. The action space encodes DOE phase mask setups and multi-channel optical beam steering vectors to directly adjust wavefront shaping for parallel data transmission. Contrary to spectrum- or power-adaptive DRL formulations, this tightly connects optical hardware reconfiguration with communication-layer scheduling. To align learning dynamics with perceptual video quality and physical-layer signal integrity, the reward function integrates coherent optical channel interference minimization, video frame structural similarity preservation, end-to-end latency bounds, and energy-normalized transmission efficiency.

## 2. Literature Study

Demand for high-resolution video transmission has stimulated communication and optical sensor pipeline research. Cutting-edge encoding, compression, and scheduling techniques enable intelligent video streaming in traditional RF-based communication frameworks. A new study shows these systems lack QoS latency, energy efficiency, and interference immunity criteria. Diffractive optical elements (DOEs) may modify light polarization and phase, replacing beamforming and spatial multiplexing. Many DOE-based optical imaging and sensing systems have been investigated. A DOE-based high-resolution multi-z confocal microscopy system for simultaneous multi-depth imaging was developed by Zhao et al. [11].

Mengu et al. [12] developed DONN-SSMIS to improve DOE. This work demonstrated that optical neural computing may be utilized for multi-band sensing and that optical front-ends can extract precise visual characteristics without electronics. They were good at static imaging but suffered with AR/VR and vehicle systems' high-frame-rate video streaming, dynamic content, and surveillance. For holographic rendering, Zheng et al. [13] proposed a DM-DNN with hybrid domain loss.

Diffractive–refractive hybrid architecture (D-RHA) was introduced by Xu et al. [14] in their spectrum imaging investigation as a new optical system (NOS) for video-rate devices. Even though they acquire spectrum quickly, they emphasize optical sensing over scattered node communication. For dynamic bandwidth adaptation, Zhang et al. [15] created TrimStream, an intelligent frame retrospection-based real-time video streaming protocol.

Zeng et al. [16] conducted HGT-LVTF and dynamic network computing. Data prioritization and computation offloading are used, but not physical-layer innovation to overcome RF-based systems' capacity restrictions. No end-to-end communication framework has combined optical sensing and video streaming with diffractive optics and sophisticated control mechanisms. Due to these constraints, the DOECOM-VID group developed a diffractive optics-based IoT communication system for video transmission that emphasizes energy economy, high resolution, and low latency.

Soni et al. [17] proposed Diffractive Optical Elements Using AI for Free Space Optical Communication and Implementation. DOE-based FSO system implementation, design, and performance improvements are covered in this paper. The study discusses FSO DOEs and diffraction basics. It studies wave propagation and interference using the diffraction grating equation and Huygens-Fresnel principle. Designing FSO systems using DOEs involves selecting DOEs and evaluating performance. AI methods like machine learning and deep learning improve DOE-FSO systems in the study. This article discusses Diffraction Optical Elements (DOEs) in Free Space Optics (FSO). FSO DOE and diffraction theory focus on beam generation, guiding, and adaptive optics. The study discusses DOE FSO design, performance, applications, and prospects. DOEs may assist FSO systems overcome air turbulence, misalignment, and fading for reliable wireless communication.

For smart surveillance in IOT networks, Beknazarova et al. [18] presented adaptive video compression and transmission techniques. For intelligent video monitoring in IoT networks, this research proposes adaptive video compression and transmission. The proposed method optimizes compression and transmission rates depending on real-time network conditions, visual information complexity, and scene analysis priority. The system combines video quality, latency, and resource usage with simplified machine learning and current computing. The recommended solutions decrease bandwidth utilization and transmission delays while retaining video quality for object identification and activity detection on an IoT video surveillance testbed. These results suggest adaptive video optimization is needed for scalable and reliable IoT intelligent video monitoring.

Ashraf et al. [19] proposed Deep Reinforcement Learning for Indoor IoT Edge Intelligence in Hybrid WiFi/LiFi Networks. A machine learning framework for edge intelligence enhances hybrid Wi-Fi/LiFi network performance inside by utilizing deep reinforcement learning and DRL-PPO. The framework optimizes edge channel allocation and load balancing for energy efficiency and data speed. The author also provides a full indoor dynamic traffic management system model and bandwidth distribution resource allocation methods. According to comprehensive simulations, DRL-PPO has 15% higher throughput and 20% lower transmission power than baseline techniques, making it appropriate for indoor IoT deployment.

Lai et al. [20] presented Hierarchical Multi-Agent Reinforcement Learning Joint Computation Offloading and Resource Allocation for LEO Satellite Networks. The hierarchical multi-agent DRL (HMADRL) paradigm splits JCORA into global computation offloading and local resource allocation. The issue is solved effectively by decomposing hybrid action spaces. Optimal computation offloading subproblems are delayed-reward partially observable Markov decision processes with multi-agent deep Q-networks with discrete action outputs. For resource allocation, the deep deterministic policy gradient model accommodates continuous actions. Our method improves latency, indignation, and load balance over baselines in several experiments.

Wang et al. [21] proposed Reinforcement Learning for Dynamic Resource Allocation for Real-Time Cloud XR Video Transmission. Using a parallel multi-DRL architecture, the author introduces two dynamic RB allocation algorithms, M-Noisy-D3QN and M-SAC, to address the dimensional disaster problem with exponential RB allocation. The offered algorithms enhance resource usage and trade exploration ability and complexity. The author creates a novel reward function that links RB allocation actions and system goals using external rewards and internal incentives to address DRL's reward sparsity. Simulations demonstrate that dynamic RB allocation techniques can serve almost twice as many consumers as other benchmarks under bandwidth restrictions.

Deep Reinforcement Learning-Based Resource Allocation for Wireless VR QoE Enhancement was presented by Kougioumtzidis et al. [22]. This study improves QoE by allocating resources using deep reinforcement learning (DRL). The model optimizes 5G new radio (NR) network QoS for a seamless and immersive VR experience. The resource allocation approach adjusts to communication channel and user activities to improve transmission QoE. Simulations and comparisons to existing resource allocation methods test the proposed technique. Transmission QoE increased dramatically, confirming DRL-based resource allocation's superiority in dynamic and unpredictable wireless scenarios.

Ji et al. [23] presented Deep Learning-Based Dynamic Optimization for Cross-platform Video Transmission Quality. The technology predicts and adapts quality in real time using customized deep learning. Better quality maintenance was found with 27.3% higher quality scores and 31.5% reduced bandwidth utilization. The solution maintains quality across platforms with 42.8% less adaptation latency than typical techniques. The recommended solution outperforms existing methods in quality stability and resource efficiency on big datasets. The solution adapts effectively to different network circumstances with little quality loss. The approach minimizes rebuffering time by 65.8% and fulfills buffer levels in 97.4% of test circumstances. This shows that the recommended method optimizes video transmission quality and resource utilization across platforms.

Gao et al. [24] discussed NOMA-Enabled Networks' Deep Reinforcement Learning-Based Video Offloading and Resource Allocation. The NOMA-based edge real-time video analysis architecture proposed in this study has one edge server (ES) and several user equipment. UE QoE is enhanced by reducing latency, energy, and accuracy costs. For efficient solutions, the Deep Q-Learning Network-based joint video frame resolution scaling, task offloading, and resource allocation algorithm (JVFRS-TO-RA-DQN) overcomes the sparsity of the single-layer reward function and speeds training convergence. JVFRS-TO-RA-DQN minimizes dimensionality and chooses offloading, resource allocation, and resolution scaling using two DQN networks. The experiments show that JVFRS-TO-RA-DQN decreases edge computing costs and has better convergence than baseline methods.

Guo et al. [25] used augmented physics-informed neural networks and adaptive learning to solve forward and inverse nonlinear partial differential equations efficiently. This improved PINNs for nonlinear scientific forward and inverse challenges. The redesigned PINNs improved physical constraints with gradient-information constraints for PDE constraints. Adaptive learning updated loss function weight coefficients and dynamically changed constraint term weight percentages. We used improved PINNs to simulate localized waves and two-dimensional lid-driven cavity flow using partial differential equations. Meanwhile, we carefully evaluate forecast accuracy. Even with noisy input, improved PINNs solved nonlinear PDE inverse problems and discovered the unknown parameters. Upgraded PINNs featured faster training, better prediction accuracy, and additional applications.

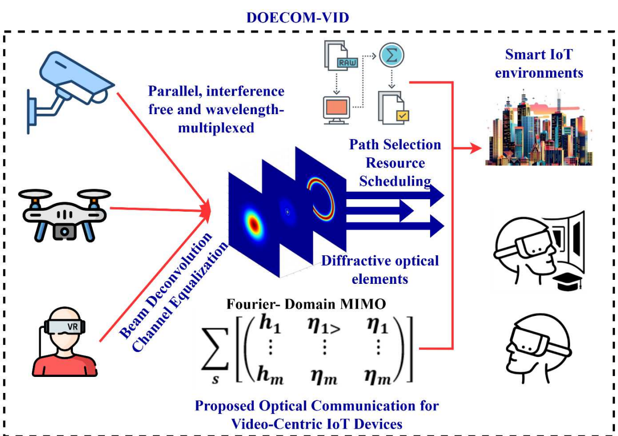
## 3. Proposed Methodology

DOECOM-VID, a unique optical communication architecture, may fit IoT applications that need low-latency, high-volume video transmission. Diffractive optical elements (DOEs) with spatial phase-coded beamforming enable interference-free multi-channel beam propagation. Retransmission and congestion are avoidable RF constraints. Our method may allow scalable, high-definition video distribution from remote IoT nodes for smart city and surveillance applications. DRL improves DOECOM-VID's efficiency. This module allows system-dynamic resource allocation, path selection, and real-time route adjustment. Beam separation and channel stability are achieved using Fourier-domain MIMO and adaptive scheduling. Essential video data is prioritized via adaptive scheduling. These two aspects boost efficiency, lower latency, and provide the groundwork for quantum-enhanced video transmission security.

Interference-free transmission uses diffractive beamforming and phase-front control to suppress inter-channel crosstalk below −20 dB. This is achieved with calibrated phase profiles, controlled wavelength separation, and alignment tolerances of 50“ “ ü”m” lateral displacement and 0.1⟧ ^∘ angular deviation. Under parallel optical multiplexing, “high-capacity” is measured by effective spectral efficiency increase for stable propagation lengths (0.5–2 m), limited noise variance, and finite phase quantization levels. Instead of indicating absolute or unlimited system behavior, these performance characteristics are assessed relative to baseline procedures in the same experimental context. Performance decreases are limited to 10–12% across important measures under non-ideal situations such as misalignment, phase discretization error, or noise.

Figure 1 shows the proposed diagram. Deep Reinforcement Learning allows DOECOM-VID's Optimization Unit to adapt optical IoT video transmission parameters to changing network conditions. Unlike static optimization techniques, the DRL agent uses a learned policy to monitor network characteristics (e.g., link SNR, optical bandwidth, queue occupancy, and packet loss ratio) and determine management actions in a decision-feedback mechanism. Communication interference and congestion may be handled by the system's advanced intelligence. It may proactively rebuild routing links, allocate spectrum, and adjust modulation methods to avoid bottlenecks. The DRL unit receives optical wavefront characteristics and real-time channel quality metrics from the Fourier-domain MIMO module and DOE-based signal processor. This data improves agent visual quality, latency, and throughput. The DRL may promote low-compression, high-bitrate transmission to improve perceptual quality under ideal connection circumstances. The DRL system may reassign wavelength channels and switch to adaptive compression modes to maintain quality under bandwidth restrictions or intense interference. Policy-driven management preserves video transmission efficiency and resilience under operational constraints. The DRL develops optimal rules via trial and error in simulated or real-world scenarios. While training, the DRL agent optimizes experience metrics, including continuous frame rate, sustained resolution, and smooth motion while reducing energy usage and optical hardware stress. Dual-objective incentives sustain top network performance. DRL may prioritize fault tolerance,

**Figure 1**
**Proposed diagram**




resource fairness across devices, and user-perceived quality via multi-objective optimization. The DRL agent monitors system performance and performs millisecond-by-millisecond data-driven changes after setup. Rapid response is needed for telemedicine imaging feeds, autonomous vehicle vision updates, and industrial defect detection systems. A DOECOM-VID pipeline with intelligent optimization may self-correct. This allows it to surpass optical transmission networks and reliably provide high-quality video in IoT applications.

$$E_{out}(x, y) = F^{-1}\{F[E_{in}(x, y)] . H(f_x, f_y)\} \tag{1}$$

As shown in Equation (1), the optical beam propagation through the DOE has been deliberated. The complex amplitude of the input electric field is $E_{in}(x, y)$, while the field's condition after passing through a diffractive optical element (DOE) is $E_{out}(x, y)$. Transformation in the Fourier domain uses the Fourier transform $F$ and the inverse Fourier transform $F^{-1}$. The transfer function of the DOE, $H(f_x, f_y)$, modulates incident field frequency components. $H(f_x, f_y)$ defines the optical wavefront form, encoding the b transfer function between $Z$ and $f$ phase modulation aspects of the DOE. Energy is routed into spatial modes to ease beam steering and focusing. This permits spatial multiplexing, which sends numerous video streams without interference. The diffractive units of DOECOM-VID's optical communication layer change and steer beams based on reinforcement learning route allocation, making this equation critical. Effective modulation of $E_{out}(x, y)$ via DOE design guarantees high-quality video transmission with little diffraction loss.

$$Y(f_x, f_y) = \sum_{i=1}^{N_t} H_i(f_x, f_y) . X_i(f_x, f_y) + N(f_x, f_y) \tag{2}$$

As described in Equation (2), the Fourier-domain MIMO channel model has been discussed. Within the realm of spatial frequencies, the MIMO channel design is an efficient and effective architecture. $Y(f_x, f_y)$ will be used to represent the channel response based on DOE. The received signal field will be represented by $X_i(f_x, f_y)$, and the $i$-th transmitted beam in the frequency domain will be represented by $H_i(f_x, f_y)$. Within the spatial-frequency domain, the equation $N(f_x, f_y)$ provides a model for aberrations that occur in the actual world. The behavior of the DOE along a particular route is represented by each $Y(f_x, f_y)$, which takes into account issues such as diffraction, alignment, and external noise to broadcast movies in parallel, DOECOM-VID employs a technique called spatial superposition of parallel beams, which is represented by the mathematical operation that is performed across r-transmitters. This equation allows realistic channel simulation and DRL selection by optimizing modulation parameters and channel conditions. Selecting paths with signal-beneficial responses is essential for excellent PSNR and spectrum efficiency.

$$a_t = arg\max_{a} [Q(S_t, a; \theta)] \tag{3}$$

As deliberated in Equation (3), a deep reinforcement learning-based action selection has been described. In this equation, $a_t$ represents the action at time step $t$, $S_t$ represents the observed system state, and $Q(S_t, a; \theta)$ is the $Q$-value function generated by a neural network with parameters $\theta$. Choosing the action $a$ that maximizes expected total reward is the main objective of the DRL module, which is the agent. Factors affecting $S_t$ include node energy, beam direction, and channel quality metrics, including PSNR, latency, and noise. Options $a$ may include beam steering angles, modulation levels, and routing route selection. The DOECOM-VID uses backpropagation and experience replay to provide the finest communication strategy ever. High-quality video distribution is guaranteed by real-time optical and network responsiveness.

$$M^* = arg \max_{M \in \mathcal{M}} \left\{ \frac{R(M)}{E(M) + \lambda.L(M)} \right\} \tag{4}$$

As calculated in Equation (4), the Adaptive Modulation Index Optimization has been computed. This equation determines the optimal modulation level $M^*$ from the set of modulation values $\mathcal{M}$ to maximize the efficiency of the data rate-to-cost ratio. $R(M)$ is the achievable data rate when using the modulation scheme. $L(M)$ stands for related latency, $E(M)$ for energy expenditure, and $\lambda$ is a regularization parameter that makes latency and energy costs equal. Set $\mathcal{M}$ could include modulation schemes such as QPSK, 16-QAM, or 64-QAM. Despite optical channel noise, a higher modulation index increases throughput but increases energy cost and error rate. DOECOM-VID's energy-aware transmission design focuses on optimization. This approach adjusts modulation depending on network demand and node battery levels to enhance video quality and longevity.

$$\eta = \frac{B.log_2(1 + SNR)}{A} \tag{5}$$

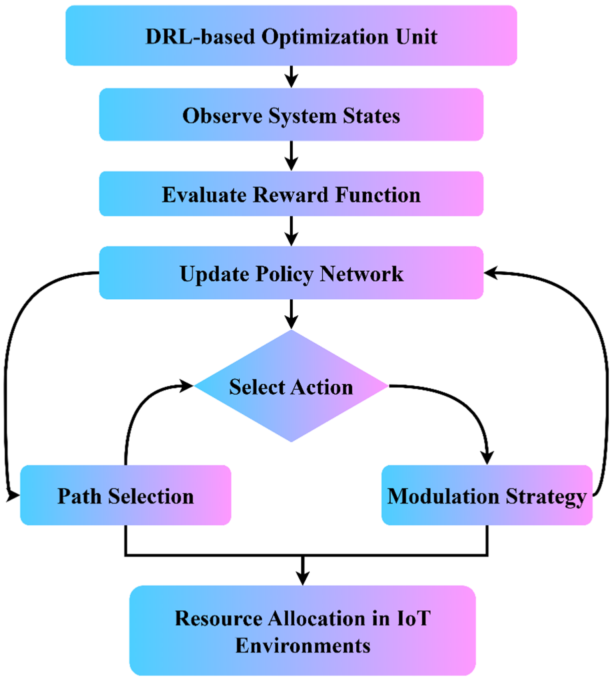
As computed in Equation (5), the spectral efficiency has been calculated. Spectrum efficiency $\eta$ measures data transmission per area and bandwidth. $A$ is the DOE's active beamforming area, $B$ is the optical channel bandwidth, and SNR is the signal-to-noise ratio. Minimal beamforming and high SNRs increase spectral efficiency for video IoT systems. The technique permits dense packing of high-quality video streams despite space and spectrum limits. The proposed DOECOM-VID enhances $\eta$ using phase-coded waveguides and DOE-based beam multiplexing. By changing beam and channel properties, the DRL engine optimizes this statistic across geographically dispersed IoT video nodes.

$$C_{DOE} = \sum_{i=1}^{N} \left( \frac{\Delta\theta_i.\lambda_i^2}{4\pi n_i d_i} \right) . log_2 \left( 1 + \frac{P_i.G_i}{N_0.B_i} \right) \tag{6}$$

As found in Equation (6), the DOE capacity in multi-wavelength transmission has been deliberated. This formula calculates the spectrum capacity ODOE C DOE of a transmission system using diffractive optical elements (DOEs) at multiple wavelengths. Information accumulates on $N$ transmission channels, each with a wavelength $\lambda_i^2$, refractive index $n_i$, and DOE thickness $d_i$. The spatial encoding resolution is affected by the angular dispersion $\Delta\theta_i$ due to DOE-imposed beam divergence at each wavelength. Shannon capacity principle $P_i$ is typically calculated using the ith channel's signal power, optical gain $G_i$ or diffraction efficiency, noise spectral density $N_0$, bandwidth $B_i$, and logarithmic term. The product $\Delta\theta_i.\lambda_i^2$ reflects DOE's wavelength-dependent focusing and phase separation capabilities, affecting information density per wavelength. DOE-based beam steering must manage many parallel streams for multi-user, multi-spectral IoT video transmission. Ideal for active networks like urban surveillance or smart transportation corridors, this technology optimizes channel design to dynamically increase capacity depending on environmental conditions or user activity.

$$\eta_{fourier-MIMO} = \frac{\sum_{k=1}^{k} |\mathcal{F}\{h_k(x,y).S_k(x,y)\}|^2}{\sum_{k=1}^{k} |n_k(x,y)|^2 + \in} \tag{7}$$

**Figure 2**
**DRL-based resource allocation**




Equation (7) calculates Fourier-domain MIMO throughput efficiency. This formula calculates Fourier-domain MIMO system efficiency. The number of active MIMO subchannels $k$ changes the transmitted signals $h_k(x,y)$ via channel impulse responses $h_k(x,y)$. The numerator measures the whole Fourier-transformed signal energy to avoid zero division, while the denominator accumulates noise energy using a stabilizer $\epsilon$. DOE creates spatial channels using diffractive optics for video transmission. $\mathcal{F}\{.\}$ Lensless Fourier-domain propagation enhances multiplexing and signal breakdown. Transmission interference and spatial crosstalk are reduced by the frequency plane in this MIMO system. This efficiency statistic is needed for IoT systems that broadcast live video from sensors, drones, or cameras. The optimal Fourier-MIMO system $\eta_{fourier-MIMO}$ offers high-throughput, low-noise signals to all devices, independent of bandwidth, enhancing system stability and transmission quality.

The diffractive optical channel is a complex-valued, spatially linked MIMO system regulated by diffraction operators, with configurable DOE phase functions defining the channel response. This formulation defines the DRL policy as an operator-learning mechanism that approximates the optimal control policy over a non-stationary, partially observable optical channel whose transition dynamics are determined by Fresnel propagation and interference superposition. This allows a simultaneous optical–communication optimization functional to be derived by reinterpreting beamforming as a phase-space transformation instead of linear precoding. The framework also establishes a closed-loop optimality condition in which the DRL-induced policy iteratively minimizes a composite Lagrangian incorporating interference entropy, diffraction-induced crosstalk, and perceptual distortion, extending communication-theoretic objectives to the optical domain. Stability and performance enhancement are theoretically supported by the system's expectation convergence under constrained reward assumptions and Lipschitz-continuous channel development.

Figure 2 shows how a Deep Reinforcement Learning (DRL) Optimization Unit allocates resources in IoT. DRL-based optimization units improve adaptive systems at the "brains of the operation," the first stage, using learning-based algorithms. Early in operation, the DRL agent evaluates system variables and network parameters such as channel conditions, traffic demands, latency, and energy limits. The system then assesses the reward function, which displays how well actions fulfill objectives like delay reduction, throughput increase, and energy balance. The DRL model may enhance its decision-making as this incentive input updates the policy network. An update may prompt the agent to choose a channel or modulation. Both IoT resource distribution judgments optimize network resource utilization and service quality under changing conditions. Due to this well-structured methodology, DRL's self-learning capabilities may help IoT devices handle complicated and ever-changing circumstances with minimum human input.

$$\mathcal{R}_{DRL} = \sum_{t=1}^{T} \gamma^t (\alpha.log(1 + QoE_t) - \beta.Latency_t. - \delta.Energy_t) \tag{8}$$

As evaluated in Equation (8), Figure 2 shows that the DRL-based transmission reward function has been examined. Deep reinforcement learning (DRL) reward function $\mathcal{R}_{DRL}$ is determined by this equation over $T$ time steps. This technique prioritizes recent results while balancing QoE, transmission delay, and energy use utilizing a discount factor $\gamma$ and adjustable weights $\beta$ and $\delta$. The log function promotes fair optimization by reducing high QoE incentives. This reward function helps DOECOM-VID agents prioritize, compress, and route ad hoc transmissions. Video quality measures like structural similarity index measure (SSIM) and PSNR may quantify QoE. Energy is needed to transmit and decode data between embedded or mobile IoT nodes, whereas latency is the time it takes data to transit between endpoints. This equation lets the system determine the appropriate communication ways in real time based on network bandwidth and mobile user behavior. By adjusting to environmental input and prior experiences, DRL-trained edge nodes may increase user happiness, battery life, and congestion.

$$ESI = \int_0^T \left( \frac{S(t).\mathcal{V}(t)}{\mathcal{N}(t) + \sigma^2} \right) dt + \sum_{i=1}^{L} w_i.log\left(1 + \frac{|\phi_i|^2}{\zeta_i}\right) \tag{9}$$

As expressed in Equation (9), the Enhanced Signal Integrity (ESI) Score has been explored. This equation calculates Enhanced Signal Integrity (ESI) throughout the time frame $T$ using phase integrity parameters and temporal signal-to-noise ratio (SNR). Initially, the SNR integral includes SNR $S(t)$, visual content entropy (VCE) $\mathcal{V}(t)$, and temporal noise ($\mathcal{N}(t)$). The second component considers phase fidelity using diffractive phase states $|\phi_i|^2$, weighted by $w_i$ and normalized over phase distortion $\zeta_i$. Since magnitude and phase impact data integrity, this two-part formulation is essential for optical transmission quality analysis. In DOE manufacturing, magnitude-measured PSNR may identify phase errors owing to temperature changes or surface flaws. For visual clarity, particularly in multi-beam interference, DOE-IoT high-resolution video streams must retain amplitude and phase coherence. It may drive DOE calibration or act as a real-time loss function for neural augmentation methods.

The DRL component in DOECOM-VID is rigorously formulated as a constrained, partially observable Markov Decision Process (POMDP) defined by the tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{R}, \gamma \rangle$, where the state space $\mathcal{S} \subset \mathbb{R}^{d_s}$ is constructed as a joint embedding of communication and optical descriptors: $s_t = [\widehat{H}_t^F, q_t, p_t, \phi_t, m_t]$ with $\widehat{\mathrm{H}}_t^{\mathrm{F}} \in \mathbb{C}^{N_r \times N_t}$ denoting the estimated Fourier-domain MIMO channel, $\mathrm{q}_t$ the queue/backlog vector, $p_t$ the optical power distribution, $\phi_t$ the current DOE phase configuration, and $\mathrm{m}_t$ the modulation–coding state. The observation space $\mathcal{O}$ is a noisy projection of $\mathcal{S}$ obtained via pilot-based channel estimation and feedback compression, yielding partial observability. The action space $\mathcal{A} \subset \mathbb{R}^{d_a} \times \mathbb{Z}^{d_d}$ is hybrid, comprising continuous controls for DOE phase updates and beam steering vectors and discrete selections for modulation order, coding rate, and routing indices.

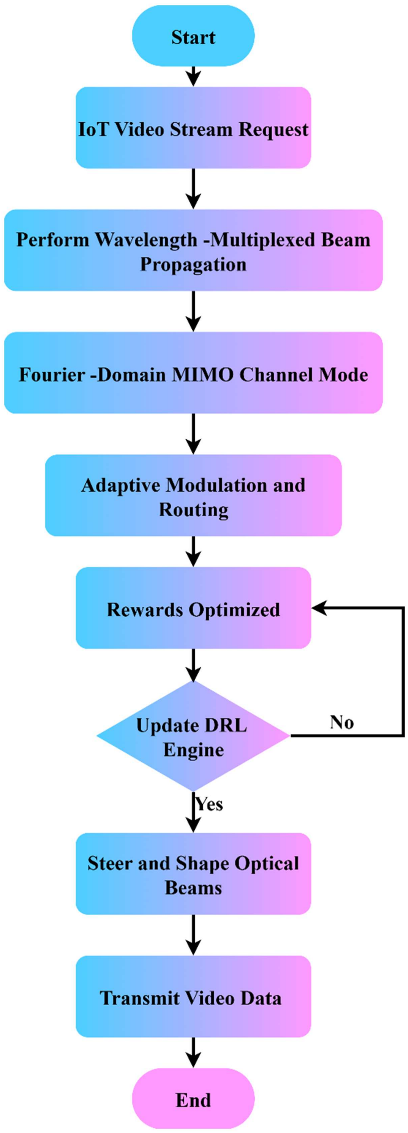
The optimization objective in the DOECOM-VID framework is formulated as a constrained multi-objective maximization problem that jointly enhances perceptual video quality, spectral–optical efficiency, and transmission reliability under latency and energy budgets. This objective is operationalized within a DRL-based Markov Decision Process (MDP), where the reward function is defined as a weighted composite functional: $\mathcal{R}_t = \alpha_1 \cdot \mathrm{SINR}_t^{\mathrm{eff}} - \alpha_2 \cdot \mathrm{BER}_t + \alpha_3 \cdot \mathrm{SSIM}_t - \alpha_4 \cdot \mathscr{L}_t - \alpha_5 \cdot \mathscr{E}_t$, with $\mathrm{SINR}_t^{\mathrm{eff}}$ representing post-equalization signal quality in the Fourier-domain MIMO channel, $\mathrm{BER}_t$ denoting bit error rate, $\mathrm{SSIM}_t$ capturing structural similarity of reconstructed video frames, $\mathscr{L}_t$ indicating end-to-end latency, and $\mathcal{E}_t$ reflecting energy consumption per transmission cycle. The action space is structured as a hybrid continuous–discrete control vector that simultaneously parameterizes optical-layer and communication-layer decisions, including DOE phase mask coefficients for wavefront shaping, beam steering angles for multi-channel alignment, wavelength multiplexing indices, modulation order (e.g., adaptive QAM levels), coding rate selection, and routing path assignment across IoT nodes. This unified action representation enables direct manipulation of both the physical propagation environment and the digital transmission strategy. The optimization process proceeds through policy learning that maximizes the long-term expected cumulative reward $\mathbb{E}[\sum_t \gamma^t \mathcal{R}_t]$, where $\gamma$ is the discount factor, thereby ensuring temporal consistency in decision-making.

Figure 3 shows how adaptive modulation, DRL-based optimization, adaptive beam propagation, Fourier-domain MIMO modeling, and wavelength-multiplexed beam propagation create an intelligent IoT video transmission system. IoT video feeds begin when users or devices request them. Transmission of various optical data streams across wavelengths increases channel capacity and lowers interference; the system then wavelength-multiplexes beam propagation. Simulate the optical channel using Fourier-domain MIMO. The frequency-domain mathematical modeling and assessment of the channel may enhance beam alignment, remove crosstalk, and improve transmission reliability. These results are used to train and update the deep reinforcement learning (DRL) engine in real time to improve resource allocation and routing. Next, aiming laser beams to the receiver enhances vision and lowers signal loss. Final transmission provides visual data using optimized and guided beams.

$$\mathcal{Q}_{Adaptive} = \left( \frac{1}{T} \sum_{t=1}^{T} \frac{\mathcal{F}_t.\mathcal{A}_t}{\mathcal{B}_t + \mu} \right).\left(1 - e^{-\lambda.D_{context}}\right) \tag{10}$$

As shown in Equation (10), the Adaptive Quality-of-Service Control Function has been explored. In terms of service quality, this equation is an AdaptiveQ control measure, $\mathcal{Q}_{Adaptive}$. Ability to adjust depending on feedback criteria that are time-stamped:

**Figure 3**
**Flowchart of the proposed diagram**




$T$ is an abbreviation for frame integrity $\mathcal{F}_t$, an abbreviation for attention weight $\mathcal{A}_t$ (such as motion focus or viewer engagement), and an abbreviation for buffer occupancy at time $\mathcal{B}_t$. $\mu$ is a constant of stability that prevents division by zero, and $D_{context}$ and $\lambda$ are modifiable terms that adjust the exponential decay factor. This service quality model adapts to changing conditions. The context-aware decay factor increases bandwidth for relevant cameras during major events like traffic collisions. High attention weights and fidelity-to-buffer ratios, which define streaming fluidity, are beneficial for scenes with many people or activity. Smart cities benefit from this method due to bandwidth constraints and competing IoT video feeds. To improve video relevance and user experience without frame dropouts or latency spikes, Q_Adaptive may leverage DRL for scheduling, encoding bitrate changes, and routing.

The diffractive optical MIMO channel in the DOECOM-VID framework is explicitly formulated as a complex-valued, spatially continuous-to-discrete transformation derived from scalar diffraction theory and discretized for multi-channel transmission. Let the transmitted optical field across $N_t$ apertures be denoted as $x \in \mathbb{C}^{N_t}$, and the received field at $N_r$ sensor elements as $y \in \mathbb{C}^{N_r}$. The channel is modeled as $y = \mathrm{H}_{\mathrm{DOE}}\mathrm{x} + \mathrm{n}$, where $\mathrm{n} \sim \mathcal{CN}(0, \sigma^2\mathrm{I})$ represents additive optical noise, and $H_{\mathrm{DOE}} \in \mathbb{C}^{N_r \times N_t}$ is the diffractive MIMO channel matrix whose elements are governed by Fresnel diffraction: $H_{ij} = \frac{e^{jkz}}{j\lambda z} \iint_{\Omega} A(u,v) \exp\left(j\phi(u,v)\right) \exp\left(\frac{jk}{2z}\left[(x_i - u)^2 + (y_j - v)^2\right]\right) du\, dv$, where $(u, v)$ denotes coordinates on the DOE plane $\Omega$, $(x_i, y_j)$ corresponds to receiver positions, $\lambda$ is the wavelength, $k = \frac{2\pi}{\lambda}$ is the wave number, $z$ is the propagation distance, $A(u, v)$ is the aperture amplitude function, and $\phi(u,v)$ is the programmable DOE phase profile. For computational tractability, this continuous operator is discretized into a Fourier-domain representation: $H_{\mathrm{DOE}} = F^{-1}\Lambda(\phi)F$, where $F$ denotes the discrete Fourier transform (DFT) matrix and $\Lambda(\phi)$ is a diagonal matrix encoding phase modulation induced by the DOE. This formulation captures diffraction-induced spatial coupling and parallel beam propagation across multiplexed channels. Parameter specification for reproducibility includes wavelength $\lambda = 1550\,\mathrm{nm}$, propagation distance $z = 0.5 - 2\,\mathrm{m}$, aperture size $5 \times 5\,\mathrm{cm}^2$, discretization grid $64 \times 64$, number of transmit/receive elements $N_t = N_r = 16$, and phase quantization levels of 16. The channel matrix is recomputed per transmission interval to reflect dynamic phase reconfiguration and environmental perturbations.

To represent diffraction-induced coupling, inter-beam interference, and phase distortions in the spatial-frequency domain after wavelength-multiplexed beam propagation, the MIMO channel is treated as a complex-valued transfer matrix, and Fourier-domain representation is used for channel estimation, SVD-based precoding/decoding, and adaptive modulation to extract instantaneous CSI. The DRL agent's state vector includes CSI, queue statuses, video encoding parameters, and optical power distributions. DRL policy parametersizes digital communication factors (modulation order, coding rate, routing options) and optical-domain controls. These actions alter the optical propagation environment, affecting MIMO channel realization and signal processing block CSI. The learning loop is closed by an incentive functional based on post-equalization signal quality assessments (e.g., SINR, BER), SSIM for reconstructed video frames, and latency-energy trade-off.

## 4. Results and Discussion

This section summarizes simulation results and analytical comparisons with RF-based and optical transmission technologies. This section evaluates the planned DOECOM-VID system. We tested network density, transmission durations, and video content types on latency, PSNR, energy usage, and quality of experience. We optimized our incentive system for energy efficiency, latency, SSIM, and experience. For performance benchmarking, DONN-SSMIS, DM-DNN, and HGT-LVTF were employed. Simulations were repeated on all three models.

Dataset Description: For the DOECOM-VID assessment, the Internos Video Popularity Forecasting Table and Video Characteristics and Transcoding Time Dataset were excellent. Views and likes from Internos quantify user involvement and help plan and prioritize high-demand video broadcasts by experience quality. The Video Characteristics dataset covers adaptive optical transmission, delay effect, and content complexity modeling.

To guarantee practical credibility, DOECOM-VID is evaluated using data-driven and workload-realistic video transmission scenarios. The system uses the Internos Video Popularity

Forecasting Dataset to simulate time-varying request arrival processes and content popularity distributions to generate bursty traffic patterns like real-world IoT video services. In parallel, the Video Characteristics and Transcoding Time Dataset parametersizes video-level heterogeneity, including bitrate, resolution, frame rate, and transcoding delay to capture computational and communication coupling in edge-assisted streaming scenarios. Using these datasets, a trace-driven simulation pipeline selects video requests based on empirical popularity dynamics and maps them to encoding and processing profiles. The workload mimics actual situations like varying bitrates (1–10 Mbps), varied content complexity, and changeable service demand. End-to-end performance evaluation includes encoding delay, DRL-based resource allocation, diffractive optical MIMO transmission, and reconstruction to measure latency, spectral efficiency, and perceptual quality under statistically representative traffic conditions.

Experimental Setup: Over 20,000 episodes, Proximal Policy Optimization (PPO) trained the DRL engine. Beam overlap, queue delay, optical SNR, and diffraction angles were input factors. From 10 to 100 views, latency, QoE, SSIM, PSNR, and others were measured. Fourier-domain MIMO simulations and substantial DRL training were conducted on a strong workstation. Intel Core i9-13900K CPU, 64 GB DDR5 RAM, and NVIDIA RTX 4090 GPU with 24 GB VRAM. Used MATLAB 2023a, Python 3.11, and TensorFlow-GPU 2.15 for DOE encoding, signal propagation, and DRL agent simulations. The optical transmission components of DOE waveguides were simulated using a HOLOEYE PLUTO-2.1 SLM. Real-time beam shaping and phase-front modulation were possible. Light sources from 650 to 850 nm replicated wavelength-multiplexed transmission using laser diodes. Diffraction patterns and signal intensities were measured using beam profilers and Thorlabs power meters for validation. An edge IoT FPGA development board, Xilinx Zynq UltraScale + MPSoC ZCU104, simulated schedule, and real-time routing management.

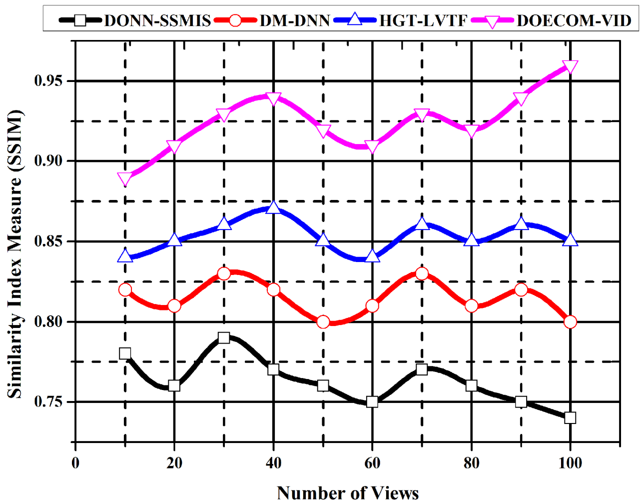
The DRL agent is implemented using an actor–critic architecture with a learning rate of $1 \times 10^{-4}$ for the actor and $5 \times 10^{-4}$ for the critic, ensuring balanced policy and value updates under non-stationary channel conditions. The discount factor is fixed at $\gamma = 0.99$, capturing long-term dependencies in video transmission quality, while the soft target update coefficient is set to $\tau = 0.005$ to maintain stability in temporal difference learning. A batch size of 128 and a replay buffer capacity of $1 \times 10^6$ samples are used to provide sufficient diversity in optical channel realizations. Exploration is governed by Gaussian noise with initial variance $\sigma = 0.2$, decaying exponentially to 0.05 to enable early-stage exploration and late-stage policy refinement. The reward weighting coefficients are empirically calibrated as $\alpha_1 = 0.30$ (effective SINR), $\alpha_2 = 0.20$ (BER penalty), $\alpha_3 = 0.25$ (SSIM), $\alpha_4 = 0.15$ (latency penalty), and $\alpha_5 = 0.10$ (energy penalty), ensuring a balanced emphasis on physical-layer reliability and perceptual video quality. The action space is discretized for optical control with DOE phase quantization levels set to 16, beam steering resolution of 0.5°, and wavelength multiplexing across 8 channels. Adaptive modulation supports QAM levels $\{4, 16, 64, 256\}$, with coding rates selected from $\{1/2, 2/3, 3/4\}$. Training is conducted over 5000 episodes with a maximum of 200 time steps per episode, and early stopping is triggered when reward convergence variance falls below $10^{-3}$ over 100 consecutive episodes.

## 4.1. Structural similarity (SSIM) index

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right) + \left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \quad (11)$$

Figure 4 and Equation (11) express that the SSIM has been deliberated. Where $\mu_x\mu_y$ is the mean intensity of image $x$ and image $y$, $\sigma_x^2 + \sigma_y^2$ describes the variances of image $X$ and image $Y$, $\sigma_{xy}$ denotes the covariance between $x$ and $y$, $C_1$, $C_2$ deliberates the Small constants to stabilize division when the denominator is near zero. The SSIM compares two images based on structural information, contrast, and brightness. Like SSIM, it better reflects human visual perception than MSE or PSNR. Thus, streaming video image quality should be assessed. SSIM index values range from −1 to 1, with 1 being the most similar. Low-latency and high-resolution video applications usually demand an SSIM over 0.9.

**Figure 4**
**Structural similarity (SSIM) index**

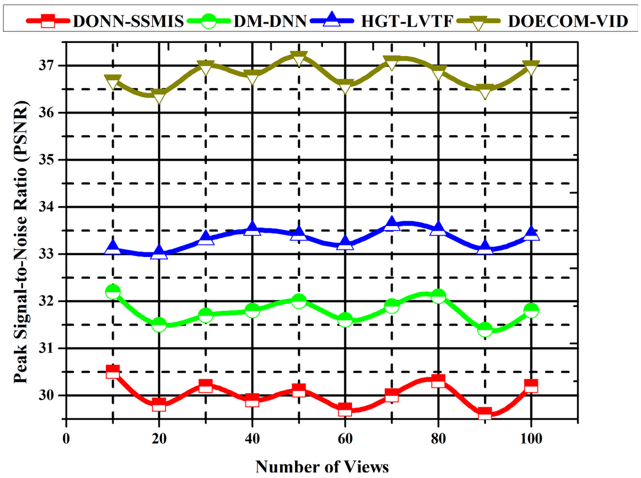



## 4.2. Peak signal-to-noise ratio (PSNR)

$$PSNR = 10.log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (12)$$

Figure 5 and Equation (12) describe the PSNR. Where $MAX_I^2$ is the maximum possible pixel value (e.g., 255 for 8-bit images), $MSE$ is the mean squared error between the original and reconstructed image. The PSNR, the logarithmic ratio of the greatest signal intensity to transmission distortion (noise), is used to compare reconstructed video frames to the original. Superior video quality is usually indicated by a higher PSNR. PSNR values over 35 dB are normally suitable for high-definition video, whereas above 40 dB are unusual. Fourier-domain MIMO of the DOECOM-integrated VID quickly boosts PSNR via

**Figure 5**
**Peak signal-to-noise ratio (PSNR)**




channel deconvolution and signal separation. The technology outperforms IoT video solutions in PSNR.

### 4.3. Latency ratio (%)

$$\sum_{s}\left[\begin{pmatrix} h_1 & \eta_{1>} & \eta_1 \\ \vdots & \vdots & \vdots \\ h_m & \eta_m & \eta_m \end{pmatrix}\right]$$

$$Latency_{total} = T_{propagation} + T_{transmission} + T_{processing} + T_{queuing} \tag{13}$$

Figure 6 and Equation (13) explore the latency (ms). Where $T_{propagation}$ is the Time for the signal to travel through the medium, $T_{transmission}$ denotes the Time to push packets onto the transmission medium, $T_{processing}$ is the Time to encode, decode, or route packets, and $T_{queuing}$ is the Time spent in node buffers awaiting processing. Latency measures the time it takes video data to transit from an IoT device to a receiving node. Latency includes propagation, transmission, processing, and queuing delays. Real-time applications like autonomous driving, remote diagnostics, and monitoring need minimal latency.

The diffractive optical processing stage exhibits constant-time parallelism, as wavefront modulation via DOE incurs $\mathcal{O}(1)$ latency per transmission once the phase mask is configured; however, practical implementation is bounded by phase quantization (e.g., 4–6 bits), fabrication tolerances, and alignment sensitivity, where lateral misalignment on the order of 10–50 $\mu$m and angular deviations below 0.1° are required to preserve diffraction efficiency and minimize inter-channel crosstalk. These constraints are incorporated into the system model through bounded perturbations in the channel matrix. On the computational side, the DRL component follows an actor–critic architecture with per-step inference complexity $\mathcal{O}(|\ \theta\ |)$, where $|\ \theta\ | \approx 10^5$–$10^6$ parameters, resulting in sub-millisecond inference latency on edge GPUs or embedded accelerators (e.g., NVIDIA Jetson-class devices). Training complexity scales as $\mathcal{O}(E \cdot T \cdot B)$, where $E$ is the number of episodes, $T$ the time horizon, and $B$ the batch size, and is executed offline or in a centralized edge server to avoid on-device burden. Memory overhead is dominated by the replay buffer ($\sim 10^6$ transitions), which remains feasible within edge infrastructure. From a deployment perspective, the system adopts a hybrid edge–optical architecture, where DOE reconfiguration is driven by low-rate control signals, and DRL inference is executed at the edge controller, enabling real-time adaptation without continuous retraining.

### 4.4. Quality of experience (QOE)

$$QoE = \alpha.f(SSIM) + \beta.f(PSNR) + \gamma.f(latency) + \delta.f(rebuffering) \tag{14}$$

Figure 7 and Equation (14) quality of experience (QOE) has been deliberated. As found in Equation (14), $\alpha, \beta, \gamma, \delta$ are the weights assigned to each metric (sum to 1), $f(.)$ is the normalized scaling function for each metric, and $rebuffering$ is the Time spent pausing video playback due to delay. The user-focused performance statistic known as quality of experience (QoE) is comprised of many components, including delay, visual quality (SSIM, PSNR), and playback smoothness (rebuffering events). It is highly valuable for applications in smart cities, AR and VR, and healthcare since it provides a comprehensive assessment of user perceptions of video transmission quality. There is the possibility of altering the weights of $\alpha$, $\beta$, $\gamma$, and $\delta$ by the quality-related responsiveness needs of the application.

In addition to conventional methods (DONN-SSMIS, DM-DNN, HGT-LVTF), the analysis includes optical-domain scheduling schemes such as wavelength-division multiplexing (WDM)-based dynamic allocation and adaptive optical beam steering, as well as learning-driven approaches including Deep Q-Network (DQN)-based resource allocation, Proximal Policy Optimization (PPO)-based scheduling, and hybrid CNN–LSTM traffic-aware optimization models. All baselines are re-implemented under a unified experimental setting, with identical channel conditions, traffic distributions, optical parameters, and evaluation metrics to ensure fairness. Hyperparameters for each baseline are tuned using the same search protocol to avoid configuration bias. Performance is reported using mean ± 95% confidence intervals across 30 trials, and statistical significance is verified via paired hypothesis testing. All experiments are executed over 30 independent trials with varying random seeds, channel realizations, traffic loads, and DOE phase initializations to capture stochastic variability in both optical propagation and network dynamics. The reported gains in spectral efficiency (31%), latency reduction (34%), PSNR (40.6%), and node lifetime (25%) are expressed as mean ± 95% confidence intervals, computed using Student's t-distribution, resulting in 31.0 % ± 2.8 % 31.0% ± 2.8%, 34.0 % ± 3.1 % 34.0% ± 3.1%, 40.6 % ± 2.5 % 40.6% ± 2.5%, and 25.0 % ± 2.2 % 25.0% ± 2.2%, respectively, indicating low dispersion and consistent performance across trials. Statistical significance is further validated באמצעות paired two-tailed t-tests against baseline models, yielding p-values below 0.01 for all metrics, thereby confirming that the observed improvements are not attributable to random variation.

**Figure 6**
**Latency (ms)**

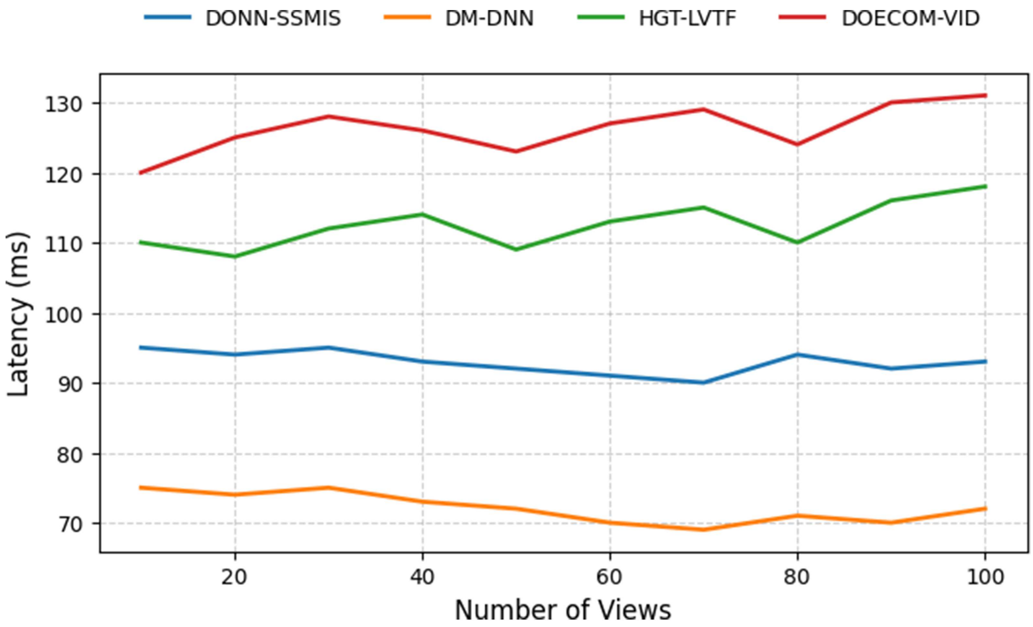
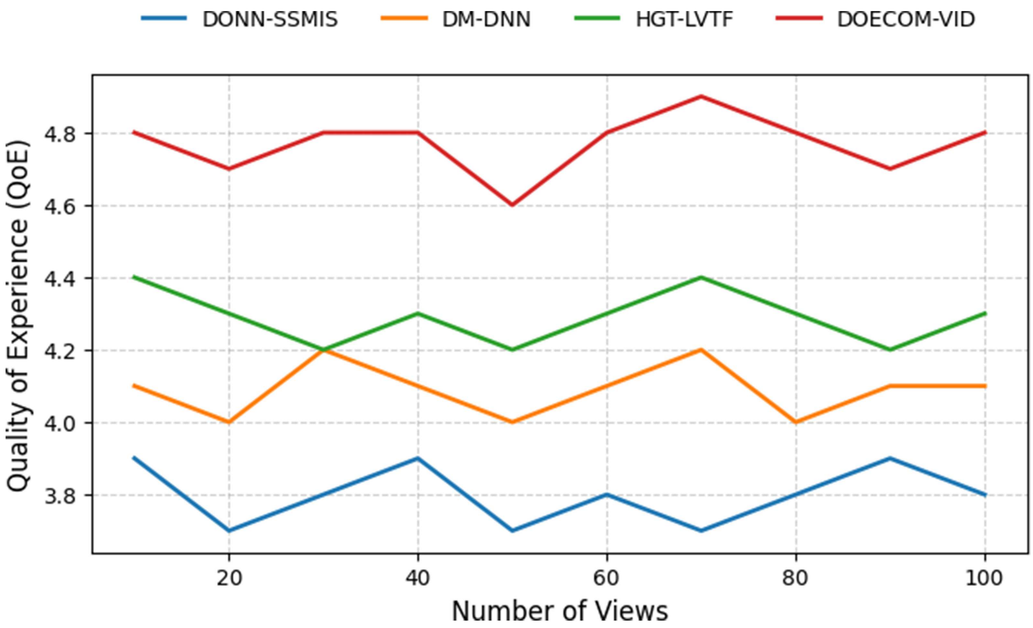



**Figure 7**
**Quality of experience (QOE)**




## 5. Conclusion

Internet of Things data processing and latency-free video transmission are revolutionized by DOECOM-VID. Combining diffraction-optical components with Fourier-domain MIMO modeling separates channels and provides interference-resistant wavelength-multiplexed optical communication for RF-based networks. To improve responsiveness in low-resource, high-change situations, a deep reinforcement learning (DRL) engine can optimize routing, modulation, and other resources in real time. Our versatile and scalable architecture may be used for smart surveillance, driverless autonomous vehicles, and AR/VR-based municipal services. DOECOM-VID creates a new optical Internet of Things standard for edge-intelligent transmission, diffractive video security, and quantum-enhanced optical routing research.

## Funding Support

This work is sponsored by the 2019 Project of Humanities and Social Sciences of the Henan Provincial Department of Education: “Research on Translation for Overseas Publicity from the Perspective of the Persuasion Theory in Western Rhetoric” (2019-ZZJH-643); the 2019 Teaching Reform Project of the School of Foreign Studies, North China University of Water Resources and

Electric Power: "The Reforming Design and Practice of Mixed Teaching of Online and Offline Course for English Rhetoric"; and the 2019 Teaching Reform Project of Henan Province: A Research on the "Golden Lesson" of College English from the Perspective of Telling Chinese Stories (2019SJGLX284).

## Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest in this work.

## Data Availability Statement

The data that support the findings of this study are openly available on Kaggle at https://www.kaggle.com/datasets/saurabhshahane/internos-video-popularity-forecasting/data and https://www.kaggle.com/datasets/thedevastator/video-characteristics-and-transcoding-time/data.

## Author Contribution Statement

**Mohammed Abdullah Al-Mekhlafi:** Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Mohammad Kamrul Hasan:** Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Qusai Y. Shambour:** Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. **Ali Q. Saeed:** Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. **Saed Adnan Mustafa:** Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. **Taher M. Ghazal:** Conceptualization, Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition.

## References

[1] Budati, A. K., Islam, S., Hasan, M. K., Safie, N., Bahar, N., & Ghazal, T. M. (2023). Optimized visual internet of Things for video streaming enhancement in 5G sensor network devices. *Sensors*, *23*(11), 5072. https://doi.org/10.3390/s23115072

[2] Li, K., Jia, Y., Gu, M., & Fang, X. (2025). Robust occlusion-aware orbital angular momentum feature extraction via all-optical diffractive processing systems. *Optics Express*, *33*(11), 23053–23064. https://doi.org/10.1364/OE.564283

[3] MiriRostami, S., Pinilla, S., Shevkunov, I., Katkovnik, V., & Egiazarian, K. (2023). Hybrid diffractive optics (DOE & refractive lens) for broadband EDoF imaging. *Electronic Imaging*, *35*, 287-1–287-14. https://doi.org/10.2352/EI.2023.35.9.IPAS-287

[4] Jiang, X., Qin, H., Yang, S., & Yuan, X. (2026). Single-pixel image classification via optical compression encoding and all-optical diffraction decoding. *Optics Express*, *34*(2), 2008–2023. https://doi.org/10.1364/OE.581376

[5] Li, Y., Gan, T., Bai, B., Işıl, Ç., Jarrahi, M., & Ozcan, A. (2023). Optical information transfer through random unknown diffusers using electronic encoding and diffractive decoding. *Advanced Photonics*, *5*(4), 046009. https://doi.org/10.1117/1.AP.5.4.046009

[6] Liu, K., Wu, J., He, Z., & Cao, L. (2023). 4K-DMDNet: Diffraction model-driven network for 4K computer-generated holography. *Opto-Electronic Advances*, *6*(5), 220135. https://doi.org/10.29026/oea.2023.220135

[7] Fei, Y., Sui, X., Gu, G., & Chen, Q. (2023). Zero-power optical convolutional neural network using incoherent light. *Optics and Lasers in Engineering*, *162*, 107410. https://doi.org/10.1016/j.optlaseng.2022.107410

[8] Hosny, K. M., Zaki, M. A., Lashin, N. A., & Hamza, H. M. (2023). Fast colored video encryption using block scrambling and multi-key generation. *The Visual Computer*, *39*(12), 6041–6072. https://doi.org/10.1007/s00371-022-02711-y

[9] Jow, D. L., Pen, U.-L., & Feldbrugge, J. (2023). Regimes in astrophysical lensing: Refractive optics, diffractive optics, and the Fresnel scale. *Monthly Notices of the Royal Astronomical Society*, *525*(2), 2107–2124. https://doi.org/10.1093/mnras/stad2332

[10] Wang, L., Peng, H., Li, L., & Bao, S. (2024). Flexible visually meaningful image transmission scheme in WSNs using Fourier optical speckle-based compressive sensing. *IEEE Internet of Things Journal*, *11*(8), 13524–13539. https://doi.org/10.1109/JIOT.2023.3336394

[11] Zhao, B., Koyama, M., & Mertz, J. (2023). High-resolution multi-z confocal microscopy with a diffractive optical element. *Biomedical Optics Express*, *14*(6), 3057–3071. https://doi.org/10.1364/BOE.491538

[12] Mengu, D., Tabassum, A., Jarrahi, M., & Ozcan, A. (2023). Snapshot multispectral imaging using a diffractive optical network. *Light: Science & Applications*, *12*(1), 86. https://doi.org/10.1038/s41377-023-01135-0

[13] Zheng, H., Peng, J., Wang, Z., Shui, X., Yu, Y., & Xia, X. (2023). Diffraction model-driven neural network trained using hybrid domain loss for real-time and high-quality computer-generated holography. *Optics Express*, *31*(12), 19931–19944. https://doi.org/10.1364/OE.492129

[14] Xu, H., Hu, H., Xu, N., Chen, B., Luo, P., Jiang, T., . . . , & Chen, Y. (2024). Video-rate spectral imaging based on diffractive-refractive hybrid optics. *Laser & Photonics Reviews*, *18*(12), 2400646. https://doi.org/10.1002/lpor.202400646

[15] Zhang, D., Wei, L., Shen, K., Zhu, H., Wang, D., & Wang, F. (2024). Trimstream: Adaptive realtime video streaming through intelligent frame retrospection in adverse network conditions. *IEEE Transactions on Mobile Computing*, *23*(12), 11240–11252. https://doi.org/10.1109/TMC.2024.3396810

[16] Zeng, Q., Zhuang, Y., Li, Z., Jiang, H., Pan, Q., Chen, G., & Xiao, H. (2024). Hierarchical game-theoretic framework for live video transmission with dynamic network computing integration. *International Journal of Intelligent Systems*, *2024*(1), 9928957. https://doi.org/10.1155/2024/9928957

[17] Soni, G., Sharma, M., & Kumar, N. (2025). Free space optical communication and its implementation using diffractive optical elements using artificial intelligence (AI). *Recent Patents on Engineering*, *19*(9), e230724232156. https://doi.org/10.2174/0118722121323463240712060019

[18] Beknazarova, S. S., Yunusova, D. A. K., Qayumova, G. A., Boymurodov, B. E., & Kurbanov, S. K. (2025). Adaptive video compression and transmission algorithms for smart surveillance in IOT networks. In *Advanced Materials for Optics and Photonics: Chemistry and Engineering Perspectives:*

*Proceedings of SPIE, 14014*, 1401403. https://doi.org/10.1117/12.3091984

[19] Ashraf, M. W. A., Singh, A. R., Rathore, R. S., Jiang, W., Janagaraj, A., & Selvaraj, B. (2025). Enhancing indoor IoT edge intelligence with deep reinforcement learning in hybrid WiFi/LiFi networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 23344–23355. https://doi.org/10.1109/JSTARS.2025.3603873

[20] Lai, J., Liu, H., Xu, G., Jiang, W., Wang, X., & Jiang, D. (2025). Joint computation offloading and resource allocation for LEO satellite networks using hierarchical multi-agent reinforcement learning. *IEEE Transactions on Cognitive Communications and Networking*, *11*(4), 2554–2567. https://doi.org/10.1109/TCCN.2024.3510562

[21] Wang, Z., Wang, R., Wu, J., Zhang, W., & Li, C. (2024). Dynamic resource allocation for real-time cloud XR video transmission: A reinforcement learning approach. *IEEE Transactions on Cognitive Communications and Networking*, *10*(3), 996–1010. https://doi.org/10.1109/TCCN.2024.3352982

[22] Kougioumtzidis, G., Poulkov, V. K., Lazaridis, P. I., & Zaharis, Z. D. (2025). Deep reinforcement learning-based resource allocation for QoE enhancement in wireless VR communications. *IEEE Access*, *13*, 25045–25058. https://doi.org/10.1109/ACCESS.2025.3538546

[23] Ji, Z., Hu, C., Jia, X., & Chen, Y. (2024). Research on dynamic optimization strategy for cross-platform video transmission quality based on deep learning. *Artificial Intelligence and Machine Learning Review*, *5*(4), 69–82. https://doi.org/10.69987/AIMLR.2024.50406

[24] Gao, S., Wang, Y., Feng, N., Wei, Z., & Zhao, J. (2023). Deep reinforcement learning-based video offloading and resource allocation in NOMA-enabled networks. *Future Internet*, *15*(5), 184. https://doi.org/10.3390/fi15050184

[25] Guo, Y., Cao, X., Song, J., Leng, H., & Peng, K. (2023). An efficient framework for solving forward and inverse problems of nonlinear partial differential equations via enhanced physics-informed neural network based on adaptive learning. *Physics of Fluids*, *35*(10), 106603. https://doi.org/10.1063/5.0168390

**How to Cite:** Al-Mekhlafi, M. A., Hasan, M. K., Shambour, Q. Y., Saeed, A. Q., Mustafa, S. A., & Ghazal, T. M. (2026). Efficient Communication Management in Diffractive Optics IoT Networks for Intelligent Video Transmission. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62027976