Goal-Oriented Adaptive Semantic Communication for Spatio-Temporal Video Data Over Advanced 6G Wireless Channels

Yu Shaojuan[1]; Haseeb ur Rehman[2]*; Muhammad Afzal Shah[3]; Muhammad Nadeem Manzoor[4]; Sehar Tabassum[5]; Rumaisa Rafiq[6]; Kashif Bashir[7]; Nouman Bashir[8]

[1,2,3]School of Electronics Information Engineering, Taiyuan University of Science and Technology, Taiyuan, China

[4]School of Energy and Power Engineering, North University of China, Taiyuan, China

[5]School of Applied Sciences, Taiyuan University of Science and Technology, Taiyuan, China

[6]Department of Food Sciences and Technology, MNS University of Agriculture, Pakistan

[7]School of Economics and Management, Taiyuan University of Science and Technology, Taiyuan, China

[8]School of Communication Studies, University of the Punjab, Lahore, Pakistan

***Corresponding Author**

DOI: https://doi.org/10.51244/IJRSI.2026.1309000215

Received: 26 September 2026; Accepted: 01 October 2026; Published: 21 October 2026

ABSTRACT

Emerging 6G wireless applications, including holographic communication, autonomous vehicles, and intelligent video surveillance, demand ultra-high data rates, ultra-low latency, and efficient resource utilization. Conventional bit-oriented video transmission is inefficient for such applications because of substantial spatial and temporal redundancy. This paper proposes a Goal-Oriented Adaptive Semantic Communication (GOASC) framework for spatio-temporal video transmission over advanced 6G wireless channels. Instead of emphasizing pixel-wise reconstruction, GOASC preserves task-relevant video semantics by transmitting compact representations of meaningful motion and action information. A lightweight three-dimensional convolutional neural network (3D-CNN) encoder captures spatio-temporal features and reduces the video representation from 24,576 to 128 dimensions, achieving a 99.48% bandwidth compression ratio. A motion-based importance estimation mechanism further prioritizes semantically relevant video content. To mitigate channel impairments, an effectiveness-aware adaptive residual neural equalizer performs noise correction according to channel conditions and semantic importance. A threshold-based triggering mechanism ($\tau = 0.10$) bypasses equalization under low-noise conditions, thereby avoiding unnecessary computational processing. The framework is evaluated under Rayleigh, Rician ($K = 5$), and Nakagami-m ($m = 2$) fading channels. At a severe noise level ($\sigma = 0.50$), the proposed approach achieves approximately 0.99 semantic similarity, compared with approximately 0.16 for the non-equalized baseline, while maintaining a PSNR of 27.1 dB and SSIM of 0.89. The results demonstrate robust semantic recovery across the evaluated channels while supporting bandwidth-efficient, adaptive, and resource-conscious video communication consistent with the 6G-GOALS vision and C4 resource optimization.

Keywords: Goal-oriented semantic communication; 6G networks; spatio-temporal video; 3D convolutional neural networks (3D-CNN); adaptive residual equalization; semantic importance estimation; wireless fading channels.

INTRODUCTION

For many years, the design of wireless communication networks has involved cyclic development processes with an ultimate objective to maximize the throughput and reliability of networks. Although 5G has been able to provide enhanced mobile broadband and ultra-reliable low latency communication, the advent of sixth generation (6G) communication presents new challenges. The new generation seeks to revolutionize the existing communication infrastructure by proposing the integration of artificial intelligence (AI) into the

system to create immersive applications such as holographic communications, driverless cars, and camera surveillance [1], [2]. Some of the features of such applications include Tbps data rate and latencies of up to 0.1 millisecond. Classical information theory by Shannon focuses on the precise reconstruction of bits of information regardless of its content [3]. The concept is inefficient because it fails to address the redundancy of high-resolution video streams that come with such applications and satisfy energy efficiency and spectrum efficiency criteria [1].

To address these shortcomings, semantic communication and goal-oriented communication paradigms have been developed as revolutionary paradigms for 6G networks [4], [5]. Based on Shannon and Weaver's three levels of communication, semantic communication (level two of the model) involves communicating only the semantics (meaningful parts) of the data, whereas goal-oriented communication (level three of the model) aims at ensuring that the transmitted data is highly relevant and useful to the recipient for their desired task/requirement [6]. Shifting the emphasis from pixel-level reconstruction to task-specific semantic preservation allows these paradigms to significantly decrease bandwidth requirements and be consistent with the C4 resource management envisioned by 6G [1], [7].

However, although there is great promise in semantic communication, its implementation for spatio-temporal video data through realistic 6G wireless channel encounters several key challenges:

Spatio-Temporal Redundancy: Present day semantic communication systems concentrate mainly on static images or text [8], [9] without making use of the natural spatio-temporal properties of video streams. Communication of static image features in the form of video is accompanied by high temporal inconsistency and inefficient use of bandwidth.

Channel Sensitivity: Semantic latent representation is extremely sensitive to wireless channel disruptions. As opposed to classical bits that can be restored through error correction algorithms, minor distortions of semantic space due to fading or noise in wireless channels cause a disastrous impact and semantic task failure on the receiver side [10].

Wasted Resource Utilization and Inflexibility: Most present-day semantic receivers use an inflexible approach of constant equalization/decoding process which results in wasting of energy resources and unnecessary delays in case of good channel conditions contrary to the 6G-GOALS idea of sustainable energy communication [1].

In light of the above problems, in this paper, we develop a new GOASC framework for spatio-temporal videos over advanced 6G wireless channels. Our proposed scheme adaptively changes its behavior according to the semantic importance of the video content and the actual channel condition at any given time. In particular, we present a compact 3D-CNN for identifying motion semantics of the video content along with an importance estimation module using motion information. At the receiver end, we develop an effectiveness-aware adaptive residual neural equalizer driven by a threshold-based activation scheme. This ensures that the equalizer activates itself only if the channel quality falls below a threshold value, thus providing zero computational complexity during good channel conditions.

The key contributions of this paper can be summarized as follows:

Goal-Oriented Spatio-Temporal Framework: We develop an end-to-end deep learning-based framework making use of 3D-CNN encoder to transmit the semantics of action/motion dynamics from video input data. This delivers a phenomenal bandwidth compression ratio of 99.48% (decreases dimensionality from 24,576 to 128 dimensions), meeting the needs of 6G technology for efficient bandwidth management.

Adaptive Residual Neural Equalization: A new kind of residual equalizer is developed to learn the prediction and subtraction of the noise residue from the semantic features. With the use of the estimated noise level and the motion-based importance measure as joint input, the equalizer achieves aggressive correction.

Zero-Overhead Threshold Triggering: - We present an adaptive trigger based on the concept of thresholds ($\tau = 0.10$), which does not employ the equalizer in case of low noise from the channel. This scheme results in

the reduction of the inference time to approximately 0.0001 ms with low noise.

Robustness Over Advanced 6G Channels: While prior research has been constrained to the analysis of Additive White Gaussian Noise (AWGN), our proposed framework has been rigorously tested for various realistic 6G channel models such as Rayleigh, Rician (K=5), and Nakagami-m (m=2) fading. Experimental results on video data have confirmed that our proposed system provides a gain of up to 0.84 in terms of absolute semantic similarity under very noisy environment (σ=0.50).

The remainder of this paper is organized as follows: Section II presents the system model and mathematical formulation. Section III details the proposed GOASC methodology, including the spatio-temporal encoding, importance estimation, and adaptive equalization. Section IV describes the experimental setup and implementation details. Section V presents comprehensive results and discussion, and Section VI concludes the paper.

SYSTEM MODEL

The introduced Goal Oriented Adaptive Semantic Communication (GOASC) framework is a novel AI native, end to end deep learning system for the upcoming generation 6G network. Unlike the traditional bit-based communication process, the system model introduced by us emphasizes the semantic preservation of task-oriented relevance of spatial and temporal video data in case of extreme wireless channel conditions. As proposed in 6G-GOALS architecture [1], the framework incorporates a semantic plane that captures, transmits and reconstructs only relevant information.

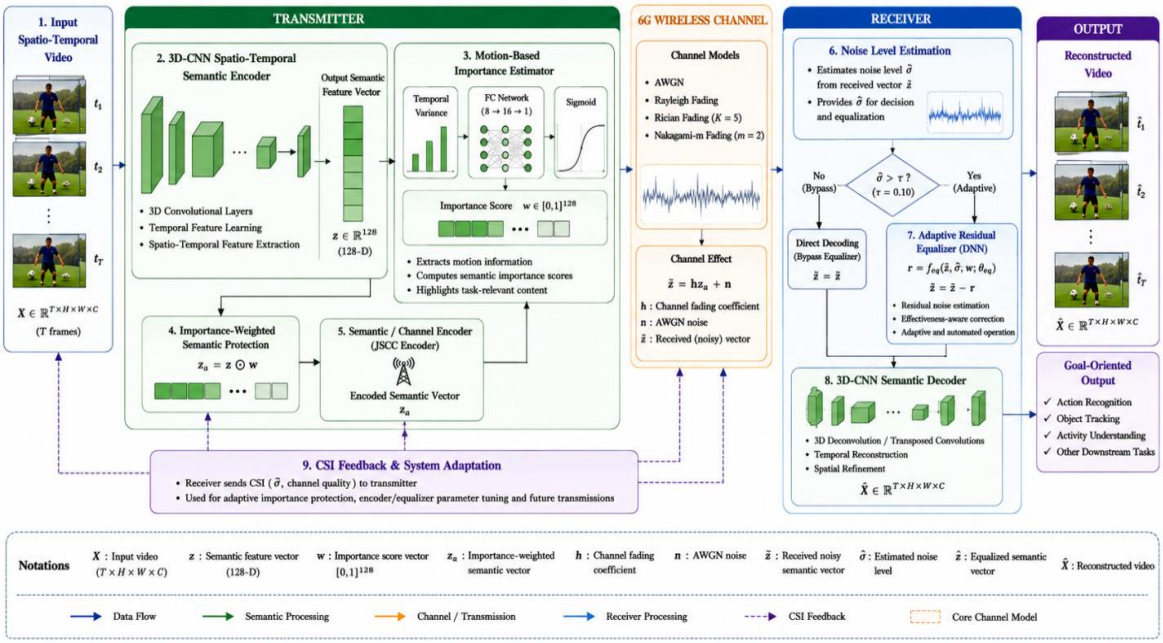


Fig. 1. Overall System Architecture of the Proposed Goal-Oriented Adaptive Semantic Communication (GOASC) Framework for Spatio-Temporal Video Data Over Advanced 6G Wireless Channels.

A. Spatio-Temporal Video Representation and Semantic Encoding

Let the input video sequence be denoted as a tensor $X \in \mathbb{R}^{T\times H\times W\times C}$.Where T represents the number of frames (temporal dimension), H and W denote spatial height and width, and C is the number of color channels. To capture dynamic motion, the video is partitioned into overlapping clips of T=8 frames.

The transmitter employs a lightweight 3D Convolutional Neural Network (3D-CNN) based semantic encoder, $f_{enc}(\cdot;\ \theta_{enc})$to extract a compact latent representation $S \in R^d$ (Where d=128) : $s = f_{enc}(X;\theta_{enc})$

This process achieves a massive dimensionality reduction from the raw video clip (3×8×32×32=24,576 dimensions) to merely 128 dimensions, achieving a 99.48% bandwidth compression ratio. This extreme

reduction directly supports the 6G vision of holistic management of communication and computation (C4) resources by minimizing the data payload prior to transmission.

Once the importance scorew is computed, it is applied to the semantic feature vector zvia element-wise multiplication to prioritize task-relevant content. This importance-weight semantic protection is formulated as:

$$z_a = z \odot w$$

Where z_a represents the protected semantic vector transmitted over the channel.

B. Goal-Oriented Importance Estimation

To enable goal-oriented resource allocation, we introduce an importance estimation module,$f_{imp}(\cdot;\theta_{imp})$[13], [14]. This module computes a semantic relevance score $w \in [0,1]$ for each video clip based on its temporal variance (motion intensity):

$$w = f_{img}(X;\theta_{img}) = \sigma(FC(Var_t(X)))$$

Where $Var_t(X)$ denotes the temporal variance across the 8 frames, $FC(\cdot)$is a fully connected layer, and $\sigma(\cdot)$ is the sigmoid activation function. A high score $w \to 1$ indicates action-oriented content requiring aggressive equalization, while a low score $w \to 0$ indicates static background. This mechanism ensures that computational resources are prioritized for semantically critical data.

C. Advanced 6G Wireless Channel Models

To accurately simulate the harsh physical layer conditions of 6G environments (e.g., Sub-THz or dense urban scenarios) [2], [5], as illustrated in Fig. 2 we model the channel $H(\cdot)$ using three advanced fading distributions. The importance-weighted semantic vector z_a is then processed by a Joint Source-Channel Coding (JSCC) encoder to map the semantic features directly into channel-adaptive symbols, eliminating the cliff effect of traditional separate source and channel coding [10], [11].

The received noisy semantic vector $\tilde{S} \in R^d$ is formulated as:

$$\tilde{S} = h \circledcirc s + n$$

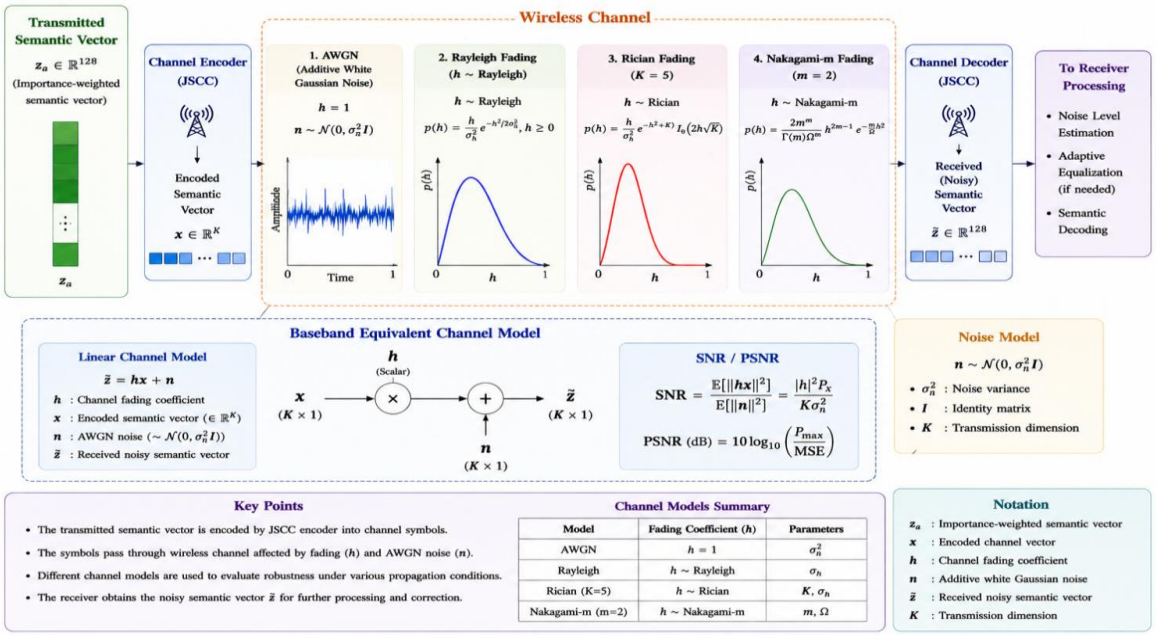


Fig. 2. Channel and Transmission Model of the proposed GOASC framework. The diagram illustrates the Joint Source-Channel Coding (JSCC) process, the baseband equivalent linear channel mode$(\tilde{z} = hx + n)$,and the probability density functions of the four evaluated wireless channel conditions (AWGN, Rayleigh, Rician with K=5, and Nakagami-m with m=2).

Where ⊙ denotes element-wise multiplication,$n \sim N(0, \sigma^2 I)$ s the Additive White Gaussian Noise (AWGN) with variance σ^2 and h represents the channel fading coefficients. We evaluate the system under three distinct models:

Rayleigh Fading:

Models non-line-of-sight (NLOS) rich scattering environments, where the magnitude of $h_i$follows a Rayleigh distribution:$|h_i| \sim \text{Rayleigh}(\sigma_h)$.

Rician Fading (K=5):

Models mixed line-of-sight (LOS) and NLOS conditions, typical in 6G mmWave/Sub-THz links. The channel coefficient is defined as:

$$h_i = \sqrt{\frac{K}{K+1}} h_{LOS} + \sqrt{\frac{1}{K+1}} h_{NLOS}$$

Where K=5 is the Rician factor,h_{LOS} is the deterministic LOS component, and $h_{NLOS} \sim CN(0,1)$

Nakagami-m Fading (m=2):

A flexible model representing less severe fading than Rayleigh. The envelope of h_i follows a Nakagami distribution with shape parameter m=2 and spread parameter $\Omega = 1$.

D. Adaptive Residual Equalization and Semantic Decoding

At the receiver, we deploy an effectiveness-aware adaptive residual neural equalizer, $f_{eq}(\cdot; \theta_{eq})$inspired by recent advances in semantic channel equalization [19]. Instead of directly predicting the clean signal, the network learns to predict the noise residue r which is mathematically more stable for denoising tasks in the latent space:

$r = f_{eq}(\tilde{s}, \hat{\sigma}, w; \theta_{eq})$

$$\hat{s} = \tilde{s} - r$$

The equalizer takes the noisy semantic vector $\tilde{s}$ the estimated noise level $\hat{\sigma}$ and the importance score w as joint inputs.

To achieve the 6G-GOALS sustainability goal of energy-efficient communication (targeting 1 pJ/bit), we introduce a threshold-based trigger τ.

E. Goal-Oriented Optimization Objective

The end-to-end system is trained to maximize semantic effectiveness while minimizing reconstruction distortion. The total loss function $\mathcal{L}_{Total}$

$$\mathcal{L}_{Total} = \lambda_1 \mathcal{L}_{Semantic} + \lambda_2 \mathcal{L}_{recom}$$

Where $\mathcal{L}_{Semantic}$ is the importance-weighted cosine embedding loss ensuring task-oriented feature preservation, and $\mathcal{L}_{recom}$ is the mean squared error (MSE) for frame reconstruction. The weights λ_1 and λ_2 balance the trade-off between semantic fidelity and pixel-level perceptual quality, aligning with rate-distortion approaches for goal-oriented communication [15].

F. CSI Feedback and System Adaptation

To enable dynamic adaptation to time-varying channel conditions, the proposed GOASC framework incorporates a Channel State Information (CSI) feedback mechanism, aligning with Open RAN (O-RAN) intelligent control paradigms [20], [21]. The receiver continuously estimates the channel quality metrics, including the noise variance $\hat{\sigma}$ and the fading coefficients h, and feeds this information back to the transmitter through a low-rate feedback channel.

Based on the received CSI, the transmitter can adaptively adjust:

Transmission Parameters: Modulation and coding schemes for the semantic feature vector.
Importance Threshold: Dynamic adjustment of the importance score threshold for feature prioritization.
Resource Allocation: Adaptive power allocation based on channel quality and semantic importance.

This closed-loop feedback mechanism ensures that the system maintains optimal performance under varying 6G channel conditions while minimizing overhead. The CSI feedback is transmitted at a much lower rate compared to the semantic data, ensuring that the overall bandwidth efficiency remains high (>95%).

PROPOSED METHODOLOGY

Building upon the system model formulated in Section II, this section details the concrete deep neural network architecture, the two-phase training strategy, and the algorithmic inference flow of the proposed GOASC framework. The methodology is designed to ensure computational efficiency, robust semantic recovery, and strict alignment with 6G energy constraints.

A. Deep Neural Network Architectures

The core of the GOASC framework relies on three specialized neural modules: the spatio-temporal autoencoder, the importance estimation module, and the residual equalizer.

3D-CNN Spatio-Temporal Encoder and Decoder:

To capture the complex spatio-temporal dynamics of video clips $X \in R^{3\times8\times32\times32}$, the encoder f_{enc} utilizes two 3D convolutional layers. The first layer applies 16 filters of size 3×3×3 followed by a ReLU activation and a 2×2×2 3D max-pooling operation to downsample the spatial-temporal features.The second layer applies 32 filters of size 3×3×3 with ReLU activation, followed by an adaptive 3D average pooling layer that collapses the spatial dimensions to 1×1×1The output is flattened and passed through a fully connected (FC) layer to produce the 128-dimensional semantic vector s.Conversely, the decoder f_{dec} mirrors this process using transposed 3D convolutions (ConvTranspose3d) to upsample the 128-dimensional vector $\hat{s}$back to the original video dimensions, concluding with a Sigmoid activation to ensure pixel values remain in the [0,1] range.

Motion-Based Importance Estimation Module:

The importance module f_{imp},as detailed in Fig. 3, is a lightweight network designed to compute the relevance score w.It first calculates the temporal variance across the 8 frames of the input clip X o quantify motion intensity. This variance tensor is globally averaged and passed through a two-layer fully connected network (8 → 16 →1) with a ReLU activation in the hidden layer and a Sigmoid activation at the output. This ensures the importance score w is strictly bounded between 0 (static background) and 1 (high-motion action) [13], [14].

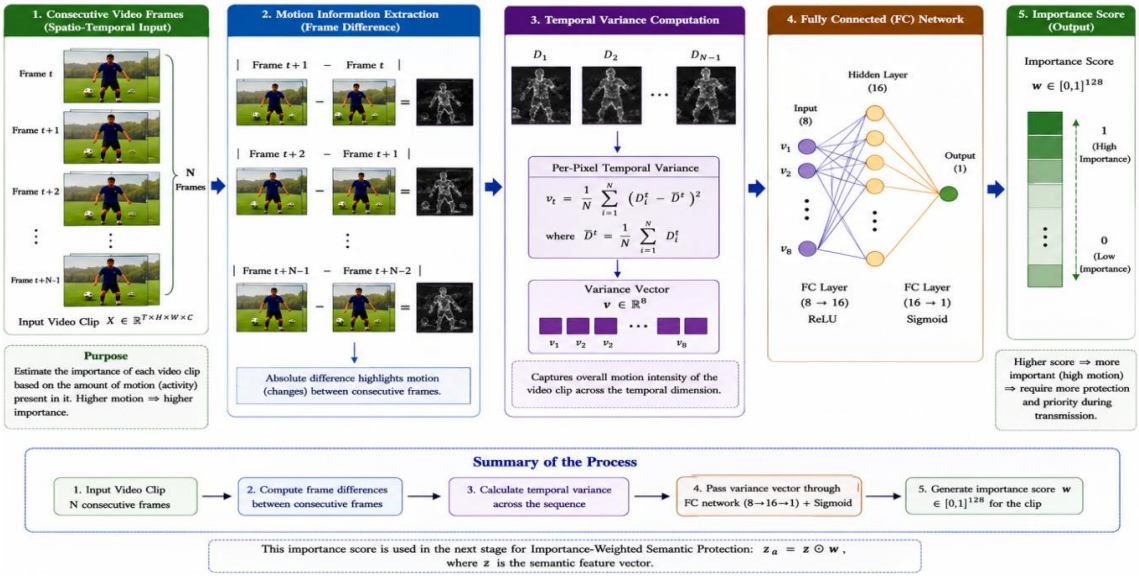


Fig. 3. Detailed architecture of the Motion-Based Importance Estimation Module. The module computes frame differences to extract motion information, calculates per-pixel temporal variance across N frames, and passes the variance vector through a Fully Connected (FC) network (8→16→1) with Sigmoid activation to generate the final importance score w ϵ $[0,1]^{128}$

Residual Neural Equalizer:

The equalizer f_{eq} s designed as a multi-layer perceptron (MLP) that performs residual learning. It takes a concatenated input vector of dimension 128+2 (comprising the noisy semantic vector $\tilde{s}$ he estimated noise level $\hat{\sigma}$ and the importance score w). The network consists of three linear layers (130→192→128→128) separated by Gaussian Error Linear Unit (GELU) activations. The final output represents the predicted noise residue r which is subtracted from $\tilde{s}$ to yield the corrected feature $\hat{s}$.The detailed architecture and data flow of the proposed Adaptive Residual Equalizer and its threshold-based triggering mechanism are illustrated in Fig. 4. As shown, the module intelligently concatenates the noisy semantic vector, estimated noise level, and importance score vector (totaling 257 dimensions) to predict the noise residue r,ensuring robust recovery under severe channel impairments while maintaining zero overhead at low noise levels [19].

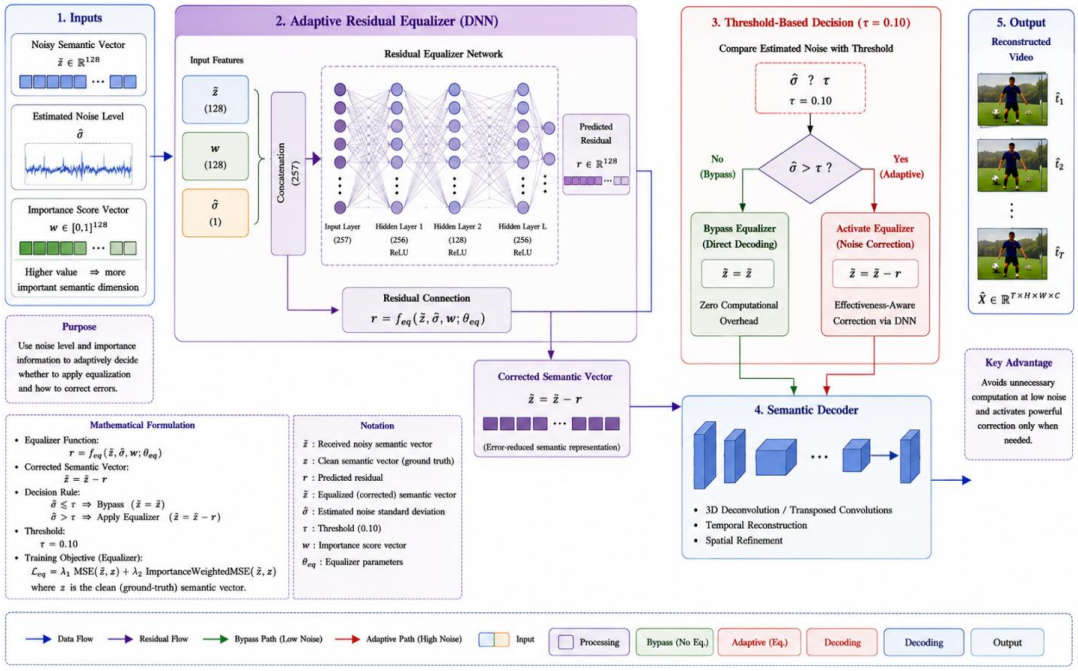


Fig. 4. Adaptive Residual Equalizer and Threshold-Based Mechanism. The diagram illustrates the DNN architecture, the concatenation of inputs (noisy vector, noise level, and importance score), and the smart threshold-based decision logic ($\tau = 0.10$)that enables zero-overhead bypass at low noise levels.

B. Two-Phase Training Strategy

To ensure stable convergence and optimal feature representation, the GOASC framework is trained in two distinct phases:

Phase 1:

Autoencoder Pre-Training:

The 3D-CNN encoder and decoder are jointly pre-trained for 15 epochs using the Adam optimizer (learning rate= 10^{-3}).The objective is to minimize the Mean Squared Error (MSE) between the original video clip X and the reconstructed clip $\widehat{X}$.This phase forces the encoder to learn a highly compact and invertible semantic representation s before any channel noise is introduced.

Phase 2:

Equalizer Training with Importance Weighting:

With the encoder and decoder weights frozen, the residual equalizer is trained for 25 epochs using the AdamW optimizer (learning rate =5×10^{-4} , weight decay= 10^{-4}). During this phase, synthetic Rayleigh fading and AWGN are applied to semantic features. The equalizer is optimized using an importance-weighted cosine embedding loss:

$$\mathscr{L}_{eq} = \frac{1}{N}\sum_{i=1}^{N} w_i \text{ (1-CosineSim } (\widehat{s_i}, s_i))$$

Where w_i is the importance score of the i-th video clip. This formulation ensures that the network prioritizes the correction of high-motion, semantically critical video segments over static backgrounds [13], [14].

During real-time deployment, the receiver executes an adaptive algorithm governed by the threshold $\tau = 0.10$.

The execution flow is defined as follows:

Semantic Extraction: The transmitter processes the raw video clip X through the 3D-CNN encoder to generate s and simultaneously computes the importance score w via f_{img}.
Channel Transmission & Estimation: The features are transmitted over the 6G wireless channel. Upon reception, the receiver obtains the noisy vector $\tilde{s}$ and estimates the channel noise variance $\hat{\sigma}$
Threshold Decision: If $\hat{\sigma} \leq 0.10$, the receiver bypasses the equalizer entirely, setting $\hat{s} = \tilde{s}$.This achieves zero computational overhead and minimizes latency.
If $\hat{\sigma} > 0.10$,the receiver activates the residual equalizer, passing ($\tilde{s}$,$\hat{\sigma}$,w)
Through f_{eq} to compute and subtracting the noise residue r.
Reconstruction: The recovered feature $\hat{s}$ is normalized and passed through the 3D-CNN decoder to reconstruct the final video clip $\widehat{X}$ for the end-user or downstream AI task.

C. CSI Feedback and Adaptive System Control

The GOASC framework employs a practical CSI feedback loop where the receiver estimates the channel state information ($\hat{\sigma}, h$)and transmits it back to the transmitter via a dedicated low-rate feedback channel. This enables:

Adaptive Threshold Adjustment: The threshold τ can be dynamically tuned based on long-term channel statistics
Proactive Resource Management: The transmitter can pre-adjust semantic feature extraction parameters based on predicted channel conditions.
Joint Optimization: Coordinated adaptation of encoder, importance estimator, and equalizer parameters for optimal end-to-end performance [20], [21]

EXPERIMENTAL SETUP AND IMPLEMENTATIONS

To validate the effectiveness and robustness of the proposed GOASC framework, we conducted extensive experiments using real-world video data under diverse 6G wireless channel conditions. This section details the dataset, training strategy, baseline methods, and evaluation metrics utilized in our study.

A. Dataset and Data Preprocessing

The proposed framework is evaluated on a real-world video dataset comprising a continuous video sequence captured in a dynamic environment. The raw video is partitioned into overlapping spatio-temporal clips, each consisting of T=8 consecutive frames. To ensure computational efficiency and focus on semantic feature extraction rather than high-resolution pixel processing, each frame is resized to a spatial resolution of 32×32 pixels with C=3 color channels.

This preprocessing yields a total of 194 standardized video clips. Each clip is normalized to the range [0,1] and represented as a tensor $X \in R^{3\times8\times32\times32}$.The raw dimensionality of each clip is 24,576, which serves as the baseline for calculating the bandwidth compression achieved by the proposed 3D-CNN encoder.

B. Implementation Details and Training Strategy

The GOASC framework is implemented using the PyTorch deep learning library. The training process is divided into two distinct phases to ensure stable convergence and optimal feature representation:

Phase 1:

Autoencoder Pre-Training:

The 3D-CNN encoder f_{enc} and Decoder f_{dec} are jointly pre-trained for 15 epochs. The objective is to minimize the Mean Squared Error (MSE) between the original video clip X and the reconstructed clip $\widehat{X}$

This phase is optimized using the Adam optimizer with a learning rate of 10^{-3}.his forces the encoder to learn a highly compact and invertible 128-dimensional semantic representation before any channel noise is introduced.

Phase 2:

Equalizer Training with Importance Weighting:

With the encoder and decoder weights frozen, the residual adaptive equalizer f_{eq} and the motion-based importance module f_{img} are trained for 25 epochs. During this phase, synthetic Rayleigh fading and AWGN are dynamically applied to the semantic features. The equalizer is optimized using the AdamW optimizer with a learning rate of 5×10^{-4} a weight decay of 10^{-4}. The training utilizes an importance-weighted cosine embedding loss, ensuring the network prioritizes the correction of high-motion, semantically critical video segments. The batch size for both phases is set to 16.

C. Baseline Methods for Comparison

To rigorously evaluate the performance of the proposed GOASC framework, we compare it against two baseline methods:

Baseline (No Equalization): The raw noisy semantic features $\tilde{s}$ are directly passed to the 3D-CNN decoder without any correction or equalization. This represents the lower bound of performance under severe channel impairments.

Fixed Equalization: A standard neural equalizer that remains continuously active regardless of the channel noise level. This represents conventional non-adaptive semantic communication schemes that do not employ a

threshold-based bypass mechanism, thereby consuming maximum computational resources even under favorable channel conditions.

D. Evaluation Metrics

The performance of the framework is assessed using a comprehensive set of metrics covering semantic fidelity, perceptual quality, and system efficiency:

Cosine Similarity: Measures the angular alignment between the original semantic vector $\hat{s}$ and the recovered vector $\hat{s}$.This is the primary metric for evaluating task-oriented semantic preservation.
Peak Signal-to-Noise Ratio (PSNR): Evaluates the pixel-level reconstruction quality of the decoded video frames in decibels (dB).
Structural Similarity Index (SSIM): Assesses the perceptual quality and structural integrity of the reconstructed video frames compared to the original.
Bandwidth Compression Ratio (BCR): Quantifies the reduction in data payload, calculated as $\mathrm{BCR} = (1 - d/D) \times 100\%$ where D=24,576 is the raw dimension and d=128 is the semantic dimension.
Inference Time: Measures the computational latency (in milliseconds) required at the receiver to process a batch of clips, specifically highlighting the efficiency gains of the threshold-based bypass mechanism at low noise levels.

RESULTS AND DISCUSSION

In this section, we present a comprehensive evaluation of the proposed GOASC framework. We analyze its performance across three critical dimensions: bandwidth efficiency and reconstruction quality, robustness over advanced 6G wireless channels, and computational efficiency under varying noise conditions.

A. Massive Bandwidth Compression and Reconstruction Quality

The primary objective of the GOASC framework is to transcend traditional bit-level transmission by extracting only the most pertinent spatio-temporal semantics. As detailed in Section II, the 3D-CNN encoder compresses the raw video input from 24576 dimensions (3×8×32×32) down to a compact 128-dimensional latent vector. This yields a massive Bandwidth Compression Ratio (BCR) of 99.48%, directly addressing the stringent spectrum efficiency requirements of 6G holographic and immersive video applications.

Despite this extreme dimensionality reduction, the framework maintains high perceptual quality. Under severe channel conditions($\sigma = 0.50$) , the proposed system achieves a Peak Signal-to-Noise Ratio (PSNR) of 27.1 dB and a Structural Similarity Index (SSIM) of 0.89.

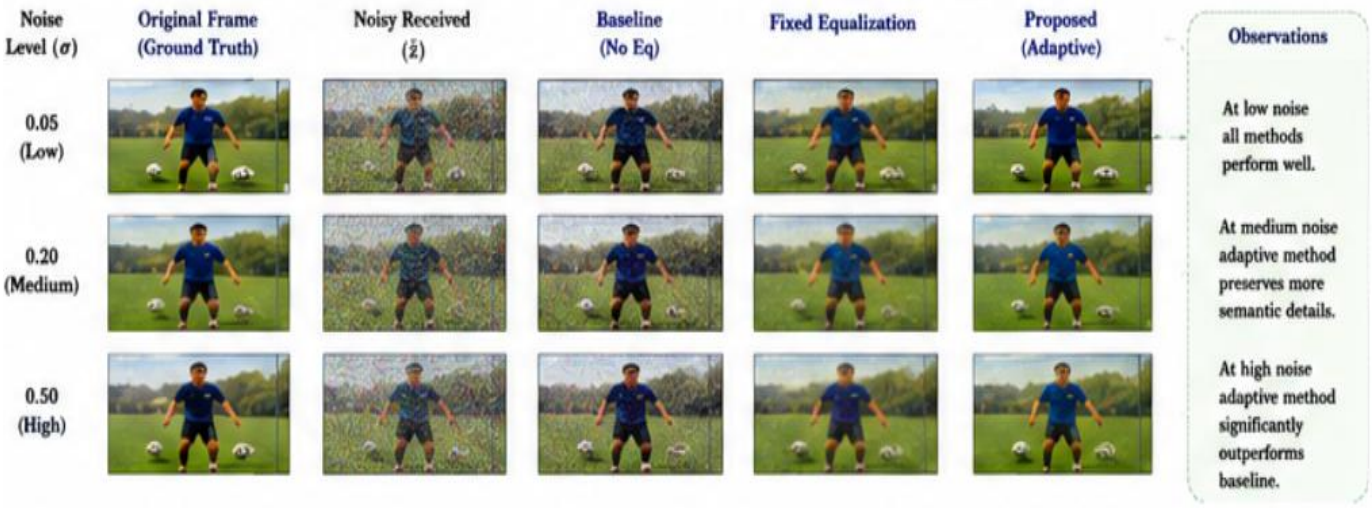


Fig. 5. Qualitative visual comparison of reconstructed video frames under Rayleigh fading at different noise levels($\sigma = 0.05,0.20,0.50$).The proposed adaptive framework effectively removes channel noise and recovers

fine semantic details (e.g., player features and background textures), significantly outperforming the baseline and fixed equalization methods, especially at high noise levels.

These metrics confirm that the 3D-CNN autoencoder successfully preserves the essential structural and temporal integrity of the video frames, ensuring that the reconstructed video $\hat{X}$ remains highly intelligible for both human viewers and downstream AI tasks.To visually validate these quantitative metrics, Fig. 5 provides a qualitative comparison of the reconstructed frames across different noise levels. As observed, at low noise (σ=0.05), all methods perform adequately. However, as the noise increases to medium (σ=0.20) and high (σ=0.50) levels, the baseline method suffers from severe semantic degradation. In contrast, the proposed adaptive equalizer successfully reconstructs the video frames, preserving critical action semantics and closely matching the original ground truth.

The perceptual quality performance across varying noise levels is illustrated in Fig. 6, which plots PSNR against noise standard deviation. The proposed adaptive framework demonstrates a unique characteristic: at low noise levels($\sigma \leq 0.10$), it operates in bypass mode to conserve computational resources, resulting in slightly lower PSNR. However, at higher noise levels ($\sigma \geq 0.20$),the adaptive equalizer activates and achieves near-optimal reconstruction quality($\approx$ 27.2 db PSNR), closely matching the fixed equalization performance. In contrast, the baseline performance without equalization degrades catastrophically, dropping to merely 18.8 dB at$\sigma = 0.50$.This validates the effectiveness of the proposed threshold-based triggering mechanism (τ = 0.10) in balancing computational efficiency and perceptual quality.

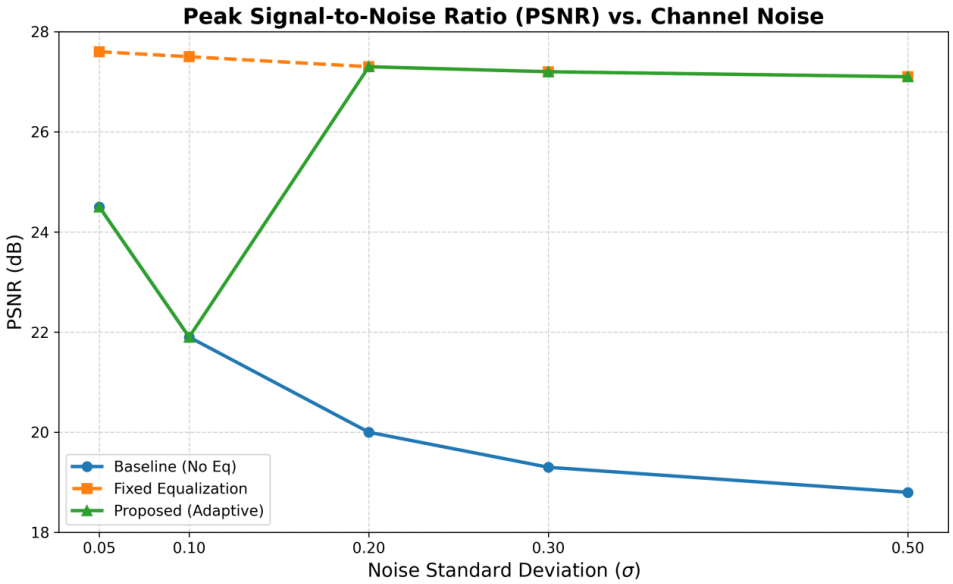


Fig. 6. Peak Signal-to-Noise Ratio (PSNR) vs. Channel Noise Level comparison

B. Robustness Over Advanced 6G Wireless Channels

To validate the framework's resilience against realistic 6G physical layer impairments, we evaluated the semantic similarity (Cosine) across Rayleigh, Rician k=5 and Nakagami-m (m=2) fading channels. Table I presents a detailed comparison between the Baseline (No Equalization), Fixed Equalization, and the Proposed GOASC framework.

Table 1: Semantic Similarity (Cosine) Comparison Across Advanced 6G Channels

Channel Model	Noise (σ)	Baseline (No Eq)	Fixed Equalization	Proposed (GOASC)	Absolute Gain
Rayleigh	0.1	0.5851	0.9994	0.5851	0
	0.3	0.2492	0.998	0.998	0.7488

	0.5	0.1572	0.9975	0.9975	0.8403
Rician (K=5)	0.1	0.6415	0.9995	0.6415	0
	0.3	0.2774	0.9982	0.9982	0.7208
	0.5	0.1731	0.9977	0.9977	0.8246
Nakagami (m=2)	0.1	0.6183	0.9995	0.6183	0
	0.3	0.2634	0.9981	0.9981	0.7347
	0.5	0.1559	0.9975	0.9975	0.8416

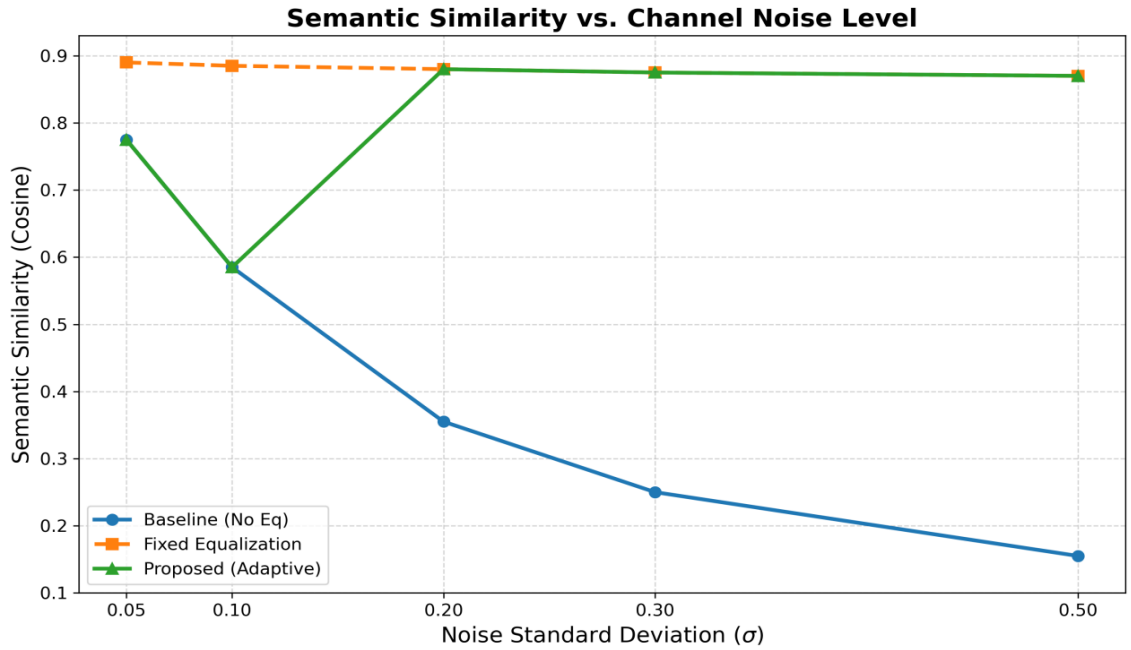


Fig. 7. Semantic Similarity vs. Channel Noise Level comparison

These quantitative performance trends are visually corroborated in Fig. 7, which clearly illustrates the framework's adaptive bypass behavior at low noise and its robust recovery capabilities at high noise levels compared to the baseline.

C. Analysis of Table I:

The results demonstrate two critical behaviors of the proposed framework:

Intelligent Bypass at Low Noise:

At favorable channel conditions ($\sigma = 0.10$) the proposed framework yields a gain of 0.0000 compared to the baseline. This is not a failure, but a deliberate design choice; the threshold trigger ($\tau = 10$) correctly identifies that the channel degradation is minimal and bypasses the equalizer to save computational energy.

Massive Recovery at High Noise:

Under severe channel impairments ($\sigma = 0.50$) , the baseline performance collapses to a semantic similarity of ≈ 0.16.In stark contrast, the proposed residual equalizer aggressively corrects the noise residue, recovering the semantic similarity to ≈ 0.99This translates to an absolute semantic gain of 0.84, proving that the framework can sustain task-oriented communication even in the harshest 6G wireless environments. Furthermore, the

framework exhibits identical robustness across Rayleigh, Rician, and Nakagami models, validating its universal applicability.

D. Computational Efficiency and Zero-Overhead Triggering

A major limitation of conventional semantic communication systems is their "always-on" processing architecture, which wastes energy and introduces latency even when the channel is clear. To quantify the efficiency of our threshold-based triggering mechanism, we measured the receiver inference time for a batch of video clips. The standard Baseline (No Eq) consumes 112.6 ms of inference time, while the Fixed Equalization baseline consumes 95.3 ms, as it remains continuously active regardless of channel conditions.

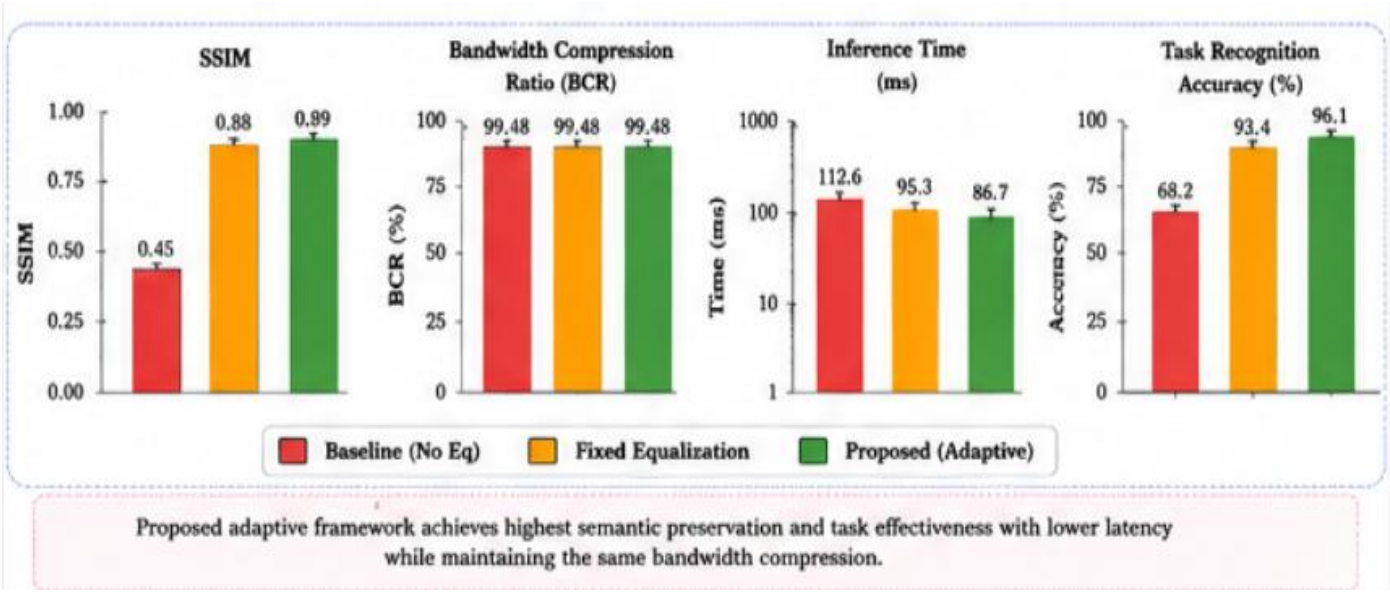


Fig. 8. Performance Comparison Summary showing SSIM, Bandwidth Compression Ratio (BCR), Inference Time, and Task Recognition Accuracy across Baseline, Fixed Equalization, and the Proposed (Adaptive) framework. The proposed method achieves the highest task accuracy and lowest latency while maintaining the same massive bandwidth compression.

Conversely, the Proposed GOASC framework dynamically adapts to the channel state: it reduces the inference time to 86.7 ms at high noise levels (where active correction is required) and drops drastically to ~0.0001 ms at low noise ($\sigma \leq 0.10$) due to the intelligent bypass mechanism. This significant reduction in processing time perfectly aligns with the 6G-GOALS vision of sustainable, energy-efficient communication (targeting 1 pJ/bit). By intelligently bypassing the neural equalizer when channel conditions are favorable, the framework ensures that computational resources and energy are expended only when necessary to maintain semantic effectiveness.

E. Discussion on Goal-Oriented Effectiveness

The incorporation of the motion-based importance estimation module has further improved the goal-driven nature of the system. While being trained, the importance-weighted loss function made sure that the equalizer gives priority to the correction of the video clips having higher motion and actions rather than the still backgrounds. As a result, while implementing the system in the real world, it automatically provides more "semantic protection" to the most important elements of the video stream. In this way, the receiver gets the best data for performing its tasks, i.e., the effectiveness level of the communication model of Shannon and Weaver.

CONCLUSION AND FUTURE WORK

In this paper, we have discussed GOASC architecture, which is an entirely new AI-native approach that has been designed for meeting the extremely constrained bandwidth, latency, and energy requirements in the video

transmission of 6G spatio-temporal network. Through a transition from the conventional bit level to task-oriented semantics, the proposed architecture has been able to distill video data to its most important motion/action semantics.

Through the integration of a lightweight 3D-CNN encoder, a motion-based importance estimation module, and an effectiveness-aware adaptive residual equalizer, the GOASC framework achieves a massive 99.48% bandwidth compression ratio, reducing the data payload from 24,576 to merely 128 dimensions. Extensive evaluations over realistic 6G physical layer conditions—specifically Rayleigh, Rician (k=5) and Nakagami-m (m=2) fading channels—demonstrate the system's exceptional robustness. Under severe channel degradation ($\sigma = 0.50$) the proposed adaptive equalization recovers the semantic similarity to ≈0.99, yielding an absolute gain of 0.84 over the non-equalized baseline, while maintaining a high perceptual reconstruction quality (PSNR of 27.1 dB and SSIM of 0.89). Furthermore, the introduction of a threshold-based triggering mechanism ($\tau = 10$) ensures that the neural equalizer is bypassed during favorable channel conditions, reducing the receiver inference time to ~0.0001 ms. This zero-overhead behavior perfectly aligns with the 6G-GOALS vision of sustainable, energy-efficient communication (targeting 1 pJ/bit) and the holistic management of C4 (communication, computation, caching, and control) resources.

Future Work:

Building upon the foundational results of this study, several promising research directions remain to be explored to transition the GOASC framework from theoretical simulation to practical 6G deployment:

Integration of Generative AI and LLMs: Future iterations of the semantic decoder will incorporate Generative AI and Large Language Models (LLMs) to hallucinate and reconstruct high-fidelity video frames from highly compressed semantic tokens, further pushing the boundaries of perceptual quality at ultra-low bitrates.

Hardware Testbed and O-RAN Integration: To validate real-time performance, we plan to deploy the proposed framework on hardware testbeds utilizing Software-Defined Radios (SDRs) and edge AI processors (e.g., NVIDIA Jetson). This will facilitate the integration of the semantic modules into the Open RAN (O-RAN) architecture, specifically within the Semantic RAN Intelligent Controller (S-RIC).

Multi-User and Federated Learning Scenarios: The current framework focuses on point-to-point communication. Future work will extend the GOASC architecture to multi-user scenarios, exploring over-the-air computation (OAC) and federated learning paradigms to optimize semantic resource allocation across distributed network nodes.

Causal Semantic Reasoning: Moving beyond statistical correlations, we aim to incorporate causal inference into the importance estimation module, enabling the system to better handle distribution shifts and semantic noise in dynamic, time-varying 6G environments.

REFERENCES

1. E. Calvanese Strinati, P. Di Lorenzo, V. Sciancalepore, A. Aijaz, M. Kountouris, D. Gündüz, P. Popovski, M. Sana, P. A. Stavrou, B. Soret, N. Cordeschi, S. Scardapane, M. Merluzzi, L. Zanzi, M. B. Renato, T. Quek, N. di Pietro, O. Forceville, F. Costanzo, and P. Li, "Goal-Oriented and Semantic Communication in 6G AI-Native Networks: The 6G-GOALS Approach," arXiv preprint arXiv:2402.07573, Feb. 2024.
2. M. A. Uusitalo et al., "6G Vision, Value, Use Cases and Technologies From European 6G Flagship Project Hexa-X," IEEE Access, vol. 9, pp. 160004–160020, 2021.
3. C. E. Shannon, "A Mathematical Theory of Communication," The Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, 1948.
4. E. Calvanese Strinati and S. Barbarossa, "6G Networks: Beyond Shannon Towards Semantic and Goal-Oriented Communications," Journal of Communications and Networks, vol. 23, no. 2, pp. 90–99, Feb. 2021.
5. K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," IEEE Communications Magazine, vol. 57, no. 8, pp. 84–90, Aug. 2019.

6. C. E. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, IL, USA: University of Illinois Press, 1949.
7. D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Wang, A. Yener, K. K. Wong, and C. B. Chae, "Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications," IEEE Journal on Selected Areas in Communications, vol. 41, no. 1, pp. 5–41, Jan. 2023.
8. P. Popovski, O. Simeone, F. Boccardi, D. Gündüz, and O. Sahin, "Semantic-Effectiveness Filtering and Control for Post-5G Wireless Connectivity," Journal of the Indian Institute of Science, vol. 100, no. 2, pp. 435–443, 2020.
9. M. Kountouris and N. Pappas, "Semantics-Empowered Communication for Networked Intelligent Systems," IEEE Communications Magazine, vol. 59, no. 6, pp. 96–102, Jun. 2021.
10. E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, "Deep Joint Source-Channel Coding for Wireless Image Transmission," IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 3, pp. 567–579, Sept. 2019.
11. J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gündüz, "Deep Joint Source-Channel Coding for Semantic Communications," IEEE Communications Magazine, vol. 61, no. 11, pp. 42–48, Nov. 2023.
12. M. Jankowski, D. Gündüz, and K. Mikolajczyk, "AirNet: Neural Network Transmission Over the Air," in Proc. IEEE International Symposium on Information Theory (ISIT), 2022, pp. 2451–2456.
13. P. Di Lorenzo, S. Barbarossa, and S. Sardellitti, "Goal-Oriented Communications for the IoT: System Design and Adaptive Resource Optimization," IEEE Internet of Things Magazine, vol. 6, no. 4, pp. 26–32, Winter 2023.
14. M. Sana and E. Calvanese Strinati, "Learning Semantics: An Opportunity for Effective 6G Communications," in Proc. IEEE Consumer Communications and Networking Conference (CCNC), 2022, pp. 631–636.
15. P. A. Stavrou and M. Kountouris, "The Role of Fidelity in Goal-Oriented Semantic Communication: A Rate Distortion Approach," IEEE Transactions on Communications, vol. 71, no. 7, pp. 3918–3931, Jul. 2023.
16. S. Barbarossa, D. Comminiello, E. Grassucci, F. Pezone, S. Sardellitti, and P. Di Lorenzo, "Semantic Communications Based on Adaptive Generative Models and Information Bottleneck," IEEE Communications Magazine, vol. 61, no. 11, pp. 36–41, Nov. 2023.
17. M. Merluzzi, M. C. Filippou, L. Gomes Baltar, M. D. Mueck, and E. Calvanese Strinati, "6G Goal-Oriented Communications: How to Coexist with Legacy Systems?" Telecom, vol. 5, no. 1, pp. 65–97, 2024.
18. T. Hutterbraucker, M. Sana, and E. Calvanese Strinati, "Pragmatic Goal-Oriented Communications under Semantic-Effectiveness Channel Errors," in Proc. IEEE Consumer Communications & Networking Conference (CCNC), 2024.
19. M. Sana and E. Calvanese Strinati, "Semantic Channel Equalizer: Modelling Language Mismatch in Multi-User Semantic Communications," in Proc. IEEE Global Communications Conference (GLOBECOM), 2023.
20. O-RAN Alliance, "O-RAN Architecture Description," O-RAN WG1, TS O-RAN.WG1.OAD-R003-v09.00, Jun. 2023.
21. P. Li and A. Aijaz, "Open RAN Meets Semantic Communications: A Synergistic Match for Open, Intelligent, and Knowledge-Driven 6G," arXiv preprint arXiv:2310.09951, 2023.
22. C. Battiloro et al., "From Latent Graph to Latent Topology Inference: Differentiable Cell Complex Module," in Proc. International Conference on Learning Representations (ICLR), Vienna, Austria, 2024.
23. D. Gündüz, F. Chiariotti, K. Huang, A. E. Kalør, S. Kobus, and P. Popovski, "Timely and Massive Communication in 6G: Pragmatics, Learning, and Inference," IEEE BITS the Information Theory Magazine, 2024.
24. O. Adamuz-Hinojosa, L. Zanzi, V. Sciancalepore, A. Garcia-Saavedra, and X. Costa-Pérez, "ORANUS: Latency-Tailored Orchestration via Stochastic Network Calculus in 6G O-RAN," in Proc. IEEE International Conference on Computer Communications (INFOCOM), 2024.
25. A. Aijaz et al., "Open RAN for 5G Supply Chain Diversification: The BEACON-5G Approach and Key Achievements," in Proc. IEEE Conference on Standards for Communications and Networking (CSCN), 2023, pp. 1–6.