Importance-Driven Semantic Resilience for Challenging Future 6G Channels

Alexander Griessel*, Cem Eteke†, Eckehard Steinbach†, Wolfgang Kellerer*
*Chair for Communication Networks, Technical University of Munich
†Chair for Media Technology, Technical University of Munich

***Abstract*—Semantic Communication has recently emerged as a novel communication strategy that prioritizes transmitting meaning over conventional bit-based transmission. By significantly reducing resource requirements in communication tasks such as video conferencing, natural language, and audio transmission, Semantic Communication promises better utilization of challenging, near radio link failure (NRLF) channels. This paper introduces a novel semantic communication framework designed to further enhance the resilience of the transmission of semantics over NRLF channels. Unlike classical, Shannon-based communication that prioritizes the perfect reception of bits, our approach focuses on ensuring the successful synthesis of the message semantics. Our framework leverages the significance of discrete semantics, and a cross-layer strategy to ensure message integrity and comprehension, even under significant loss. Key to our framework is the novel, code block based Proactive Redundancy Transmission (PRT) mechanism prioritizing critical semantics, coupled with a novel error concealment step enabling meaningful reconstruction of non-critical semantics. We establish the resulting importance-driven resilience optimization problem, and introduce and validate preliminary heuristics as an initial attempt to optimize it. We formalize, implement, and evaluate our framework, demonstrating significant improvements in the resilience of semantic communication in NRLF environments. Our evaluations, leveraging a First Order Motion Model (FOMM) for video conferencing synthesis, underscore the resilience of our semantic communication framework against traditional H.265 compression under challenging Channel Block Error Rates (CBLERs). Unlike H.265, which fails to decode under significant CBLERs, our method exhibits remarkable resilience, maintaining perceptual quality even with CBLERs surpassing 75%.**

I. Introduction

As mobile devices feature in increasing capacity in our daily routine, user expectations for high-quality, widely available services grow proportionally, irrespective of user environment. Simultaneously, more real-time human and machine data is generated than ever before at the network edge, imposing new demands on available datarates and network coverage [1]. The number of connected devices in the consumer, industrial and healthcare sectors is unlikely to peak, with their service availability constraints posing a challenge that cannot be overcome by existing 5G technology [1]–[3]. To combat increased resource congestion, the 5G NR standard introduces the mmWave FR2 spectrum. However, new bandwidth standardization is limited in both availability and service capability, as new low frequency bands capable of wide area service coverage and good building penetration face significant spectrum scarcity. These constraints have lead to the emergence of Semantic Communication (SC) as a novel communication concept, promising advanced exploitation of allocated spectrum, with the low data rate requirements of SC well covered in existing work [4]–[6]. However, we propose that SC not only promises a significantly reduced required goodput compared to classical Shannon communication, but also increases service coverage under congested channels or near radio link failure (NRLF). This is achieved through the unique approach of distilling information to its essential semantics, with the ability to prioritize the inherently discrete semantics by their significance. This independence ensures that even in the absence of the complete semantic vector, the conveyed semantics retain useful meaning [7], [8]. We propose that this characteristic of SC not only facilitates a higher resilience to symbol loss but also enhances the robustness of communication in challenging conditions, as the partial reception of semantic information still provides value and actionable insights.

While SC has been recognized for its efficiency in goodput, its potential for enhancing resilience under challenging channel conditions—particularly through robust transmission and reception of semantics despite significant loss—remains largely untapped. This resilience marks a departure from traditional communication models that emphasize the error-free delivery of data bits, pivoting instead to focus on the integrity and understanding of the message's semantics, offering a strategic advantage in maintaining communication quality and reliability. In this context, we introduce Proactive Redundancy Transmission (PRT), a novel, semantic-based strategy designed to supersede the classical Incremental Redundancy Hybrid Automatic Repeat reQuest (IR-HARQ) mechanism. PRT shifts focus from exact data recovery to prioritization and preservation of essential semantics, thereby enhancing communication resilience, potentially reducing latency and feedback overhead, and marking a significant advancement in efficient network resource utilization. Additionally, PRT exploits the independent decodability of code blocks, allowing for selective redundancy and further mitigating the risk of semantic loss by capitalizing on the low goodput requirements of SC. Our approach leverages SC's unique capacity to prioritize and transmit the most critical semantics, ensuring their preservation even in adverse conditions through a novel application-agnostic framework. This framework embeds application-specific semantic mechanisms within the application layer and employs cross-layer feedback, thereby enhancing semantic task resilience and introducing an error concealment mechanism that infers missing semantics. Our paper delves into the optimization challenges and solutions for PRT, evaluating its performance through extensive testing on a video synthesis task. This analysis confirms the efficacy of our framework, offering significant improvements in semantic communication reliability.

The remainder of this paper is organized as follows: Section II reviews the state of the art in semantic communication frameworks and resilience mechanisms. Section III introduces our proposed framework for enhancing the reliability of semantic communication tasks. Section IV details the system model and the optimization problem for our Proactive Redundancy Transmission (PRT) approach. Section V assesses our framework through a video synthesis task, showing how it performs in practice. Finally, Section VI concludes the paper and outlines future work.

II. State of the Art

For the purposes of this paper, a semantic symol s_i is defined as a discrete unit of meaning. Within the network stack, semantic symbols can be treated and handled independently. As previously discussed (see [4]–[6]), semantic representations are highly compressed, and require much lower bitrates when compared to conventional representations. While the goodput advantages of SC have been thoroughly investigated, we contend that the inherent advantage of semantic symbol extraction towards the resilience of SC has not been fully explored, despite some initial work on improving Quality of Experience (QoE). In [4], notable improvements in perceptual loss at high BERs (> 0.1) are achieved through task-specific training for video communication, which poses challenges in generalization and introduces substantial computational and latency overheads. Similarly, their SVC-HARQ approach adds complexity by requiring semantic-aware error correction and training for channel gain variations, limiting its adaptability and increasing system demands. These aspects underscore the challenges in applying SVC-HARQ broadly across varying network conditions and applications. In Sec. V, we demonstrate that similar improvements w.r.t. perceptual loss are achieved by taking advantage of the symbolization of semantics, even in a generalized framework. The proactive scheduling of radio resources is not unique to SC. In [9] and [10], resources are reserved for URLLC traffic in a joint eMBB and URLLC scenario. [11] suggest pre-scheduling retransmission slots for URLLC traffic to reduce retransmission latency.

Defining a measure of semantic importance is a critical aspect of SC, and not a novel endeavor. In [3], the concept of Data Importance Information (DII) is proposed, to enable adaptive transmission, multi-access, and resource allocation at lower layers. A base station resource block allocation policy attentive to the importance distribution of semantics across multiple users is developed in [7]. For speech applications, [12] develop a semantic encoder prioritizing speech signals with high amplitude over signals with low amplitude. A key shortcoming of existing approaches is the lack of an application-agnostic importance measure, which is necessary for the generalization of SC. Depending on the application, semantic symbols may be defined as triples [7], feature vectors [13], or even as data chunks of undefined length [14], [15]. Therefore, we contend that the task of generating these specific importance values, and their subsequent mapping to generalized importance values for the network stack, should be subject to the application's implementation. This strategy enables the network stack to allocate semantic symbols to available resources, agnostic to the application task.

III. Framework

We propose a SC framework that is capable of extracting semantic symbols from an information source, assigning importance to the semantic symbols, and transmitting the symbols over a wireless channel. At the receiver, the framework is capable of synthesizing the received meaning, and preserving the intended goal in the presence of channel errors. We design our framework to be agnostic to the application and the specific meaing being transmitted. We develop our framework on top of the existing 5G NR network stack, interfacing the semantic representation generated at the application layer with the link layer. In Fig. 1, we illustrate the system model. The framework is composed of the following components:

- **Semantic Extractor**: The semantic extractor $E(\cdot)$ is responsible for extracting semantics from an information source. The information source can be an image, video, audio, or text. The semantic extractor can be implemented using a variety of techniques, such as deep neural networks, or manually crafted functions. We envision the extraction of semantics from an information source as a set of discrete units of meaning- *symbols*.
- **Semantic Importance Assignment Function**: The semantic importance assignment function $P(\cdot)$ is responsible for assigning an importance value to each semantic symbol. The importance value is used to prioritize high-value semantics. The importance value is determined by the application, and approximates the effect of the loss of a given symbol to a loss in receiver Quality of Experience (QoE).
- **Semantic-aware Channel Encoder**: The semantic-aware encoder $C(\cdot)$ is responsible for encoding the semantic information using the assigned importance values. The encoder is capable of filtering, duplicating, and allocating resources according to symbol importance, and is composed of the allocation step C_{alloc} and the encoding step C_{enc}. For the purpose of this paper, we simplify the semantic-aware encoder to a single step, $C(\cdot)$.
- **Semantic-aware Channel Decoder**: The semantic-aware decoder $C^{-1}(\cdot)$ is responsible for decoding the received semantic information. The decoder is capable of preserving correctly decoded code blocks, even when the transport block is corrupted, and especially under conditions with low guarantee of successful retransmission. The decoder is composed of the decoding step C_{dec} and the recovery

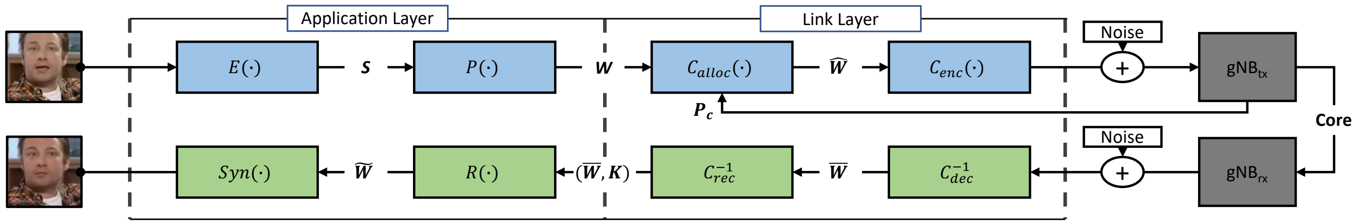

Fig. 1: System Model of the End-to-End prioritized Semantic-Aware Communication Network

step C_{rec}. For the purpose of this paper, we simplify the semantic-aware decoder to a single step, $C^{-1}(\cdot)$.

- **Semantic Error Concealer**: The semantic error concealer $R(\cdot)$ is responsible for concealing errors in the received semantic information. The error concealer is capable of utilizing inter-symbol or temporal correlation between symbols, and the inherent understanding of their semantics, to enable highly successful error concealment techniques.
- **Semantic Synthesizer**: The semantic synthesizer $Syn(\cdot)$ is responsible for synthesizing an appropriate representation of the received semantics, where the synthesis accomplishes the intended goal of the application.

Semantic symbols $S = \{s_1, s_2, \ldots, s_i\}$ are crucial for understanding and further processing semantics derived from an information source I (e.g. images, text, or audio samples) via $E(\cdot)$. In literature, Deep Neural Networks (DNNs) are commonly used to implement $E(\cdot)$ [6], [16]. However, the implementation of $E(\cdot)$, whether it is a DNN or a manually crafted function, depends on the specific requirements of the application. For example, in a video streaming application, $E(\cdot)$ may be a DNN capable of extracting the key objects and relationships in a video frame [13], [14]. In a speech recognition application, $E(\cdot)$ may be a DNN capable of extracting the phonemes and words from an audio sample. In a text summarization application, $E(\cdot)$ may be a manually crafted function capable of extracting the key sentences and phrases from a text document. Following extraction, the semantic importance of each symbol is quantified by their weights $W = \{(\rho_1, s_1), (\rho_2, s_2), \ldots, (\rho_i, s_i)\}$, assigned by $P(\cdot)$. This assignment approximates the impact of each symbol on the Quality of Experience (QoE). The subsequent encoding process is sensitive to the state of the channel P_c, which includes considerations such as grant size, code block configuration, and modulation order, ensuring that the encoding of semantic information, $\hat{W} = C(W, P_c)$, is both efficient and effective. Upon reaching the receiver, a decoding process reconstitutes the semantic symbols. Undecoded symbols are addressed by a semantic error concealer that leverages symbol history H, the decoded symbols $\bar{W}$, and the indices of undecoded symbols K, to generate a synthesized final semantic representation, $\bar{I} = Syn(\tilde{W} \cup \bar{W})$.

IV. System Model

We model an end-to-end connection established over the Uplink and Downlink channels of a 5G NR network, where the Uplink Base Station (BS_{UL}) is notified of the UE_{TX} semantic application, its resource demands, and accordingly allocates resource block grants to the UE_{TX}. Further, the BS_{UL} determines an appropriate number of code blocks, associated code block size given the scheduled grant, the needs of the UE_{TX}, and the current channel state. The UE_{TX} can then allocate one or more symbols per code block, and transmit the code blocks over the Uplink channel, where the BS_{UL} can independently decode each block and forward the correctly decoded blocks. This process incurs overhead penalties due to the additional control information required to coordinate the allocation of resources, as well as overhead due to additional appended CRCs. However, as shown in Sec. V, this leads to significant gains in performance even at high error rates, and enables resilience to highly fluctuating channels.

A. Transmission

We consider the transmission of S over a wireless channel, and the allocation of symbols to available resources. Specifically, the semantic-aware encoder $C(\cdot)$ is responsible for allocating S to the available resources in a way that maximizes the QoE at the receiver. The semantic extractor $E(\cdot)$ generates a stream of semantic *frames*, with a variable N symbols per frame, and n_i bits per symbol. Additionally, each symbol is assigned an importance weight ρ_i, by the semantic importance assignment function $P(\cdot)$. Under a constrained channel, $C(\cdot)$ must allocate symbols from each frame, such that the potential QoE at the receiver is maximized. In the face of a lossy channel, practical guarantees for the arrival of single symbols are not possible. In this regard, the goal of a transmission should be to maximize the number of high-importance symbols to arrive at the receiver. Therefore, $C(\cdot)$ must allocate symbols under both constrained *and* lossy channel conditions. With constrained resources, $C(\cdot)$ must prioritize symbols according to their importance, while under lossy channel conditions, $C(\cdot)$ must allocate symbols in order to maximize the probability of high-importance symbols arriving at the receiver. We formalize the optimization problem for the allocation of symbols under constrained and lossy channel conditions with the following parameters. As quantification of QoE is dependent on the implementation of the application, while symbol weights can be treated agnostic to the application, we define the optimization problem in terms of symbol weight.

B. Optimization Problem Formulation

In the context of SC, where each message is composed of semantic symbols with varying degrees of importance,

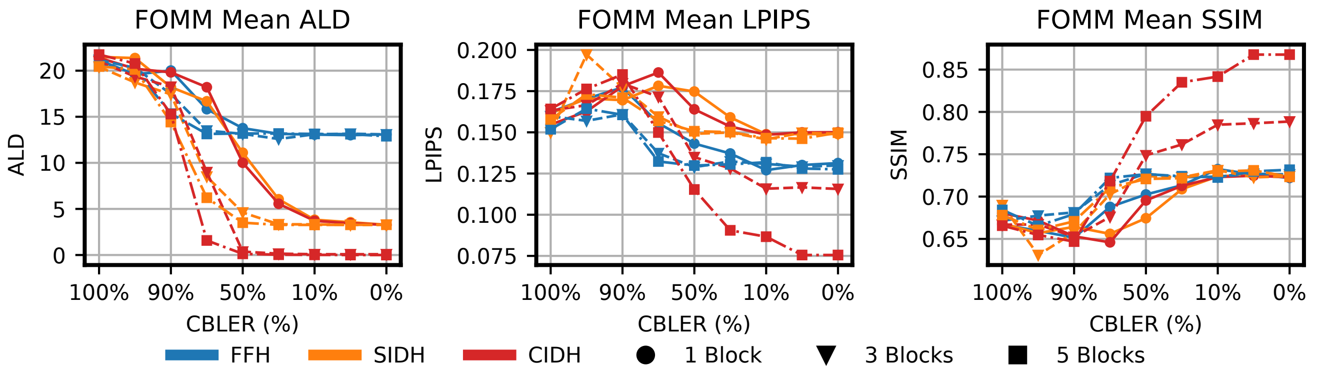

Fig. 2: ALD, **LPIPS** (lower is better), and **SSIM** (higher is better) scores for the FOMM model at mean 2 symbols per block, evaluated at varying CBLERs. The proposed CIDH strategy outperforms FFH and SIDH for all metrics. CIDH and SIDH benefit from increased block count.

we aim to optimize symbol allocation across multiple code blocks within a transmission frame. Consider a semantic frame consisting of N symbols to be transmitted over K code blocks, each with a capacity of C_j bits. Each symbol i, characterized by a length of n_i bits and an importance weight ρ_i, must be allocated to one or more code blocks to maximize the overall quality of the transmitted message. To formalize this, we introduce a binary decision variable d_{ij}, indicating whether symbol i is allocated to code block j ($d_{ij} = 1$) or not ($d_{ij} = 0$). The reception probability p_i for each symbol i is a function of the Code Block Error Rate (CBLER) and the number of symbol duplications D_i, represented as $p_i = 1 - (\text{CBLER})^{D_i}$. This formulation captures the trade-off between symbol duplication for reliability and efficient use of code block capacity.

Objective Function: Our objective is to maximize the sum of the weighted importance of symbols successfully received, accounting for the probability of each symbol's reception:

$$\max W_\rho = \sum_{i=1}^{N} \rho_i \cdot p_i \tag{1}$$

Constraints: The optimization problem is subject to the constraints:

$$\sum_{i=1}^{N} n_i \cdot d_{ij} \leq C_j, \qquad \forall j = 1, 2, \ldots, K \tag{2}$$

$$\sum_{j=1}^{K} d_{ij} \geq 1, \qquad \forall i = 1, 2, \ldots, N \tag{3}$$

$$d_{ij} \leq 1, \qquad \forall i \in N, j \in K \tag{4}$$

Where constraint Eq. 2 ensures that no code block is overloaded, Eq. 3 mandates each symbol's allocation to ensure its potential reception, and Eq. 4 prevents multiple allocations of a single symbol to the same code block (which would bring no gain in Eq. 1)

C. Symbol Allocation Heuristics

Three distinct symbol allocation heuristics are considered: the *First-Fit Heuristic* (FFH), the *Strict Importance-Driven Heuristic* (SIDH), and the *Cyclic Importance-Driven Heuristic* (CIDH). FFH is straightforward, focusing solely on semantic symbolization without considering symbol importance. SIDH prioritizes symbols based on their importance to maximize the likelihood of high-importance symbols being received. CIDH, on the other hand, balances importance with the attempt to guarantee that each symbol is sent at least once.

- **First-Fit Heuristic (FFH)**: Allocates symbols based on their appearance sequence, disregarding importance. It ensures a steady allocation rate but doesn't prioritize symbols by importance, adhering to constraints Eq. 2 and Eq. 3.
- **Strict Importance-Driven Heuristic (SIDH)**: Symbols are allocated to resources according to their importance. High-importance symbols are allocated at a higher rate, and are allocated before low-importance symbols. Further, when the code block size is insufficient to allocate all symbols, low-importance symbols remain unallocated. SIDH meets constraints Eq. 3, but does not meet constraint Eq. 2. We implement SIDH to demonstrate the importance of Eq. 2.
- **Cyclic Importance-Driven Heuristic (CIDH)**: A compromise between FFH and SIDH, CIDH allocates symbols according to their importance, but also ensures that all symbols are sent at least once, if possible. High-importance symbols are allocated at a higher rate, and are allocated before low-importance symbols. CIDH satisfies all constraints.

The effectiveness of these heuristics is assessed in Sec. V.

V. Implementation and Evaluation

The proposed framework is implemented and evaluated on a First Order Motion Model (FOMM) [13] based video synthesis task. The FOM model is trained on the VoxCeleb2 dataset [17] and is used to synthesize a head and shoulders video conference. The keypoint detector $E_{\text{FOMM}}(\cdot)$ extracts facial landmarks from a source image and a driving video, yielding a feature vector of fixed size, containing the positions and jacobians of the landmarks. The feature vector is symbolized into ten equal semantic symbols. The feature vector can then be used to synthesize a video from the contained landmark data and a target image. As a baseline for comparison of performance, we compare the FOMM model against H.265 compressed video.

TABLE I: FOMM PRT Heuristic Objective Scores

Test	K	Sym/C_j	FFH (W_ρ, μN)	SIDH (W_ρ, μN)	CIDH (W_ρ, μN)
1	1	2 / 10	0.10, 2.0	0.16, 2.0	0.16, 2.0
2	3	2 / 10	0.26, 4.9	0.23, 2.0	0.30, 6.0
3	5	2 / 10	0.38, 6.7	0.26, 2.0	0.5, 10.0
4	1	4 / 10	0.20, 4.0	0.27, 4.0	0.27, 4.0
5	3	4 / 10	0.45, 7.8	0.40, 4.0	0.53, 10.0
6	5	4 / 10	0.59, 9.2	0.44, 4.0	0.65, 10.0
7	1	9 / 10	0.45, 9.0	0.47, 9.0	0.47, 9.0
8	3	9 / 10	0.70, 9.9	0.68, 9.0	0.70, 10.0
9	5	9 / 10	0.78, 10.0	0.75, 9.0	0.78, 10.0

A. Synthesis Evaluation Metrics

The FOMM task is evaluated using the Average Landmark Distance (ALD), Learned Perceptual Image Patch Similarity (LPIPS) [18], and Structural Similarity Index Metric (SSIM) [19] metrics. The ALD metric measures the average distance between the ground truth landmarks and landmarks detected in the synthesized frame. The LPIPS metric measures the perceptual similarity between patches in the ground truth and synthesized frames. The SSIM metric measures the structural similarity between patches of the ground truth and synthesized frames. We use ALD to measure the accuracy in reproduction of facial landmarks in the synthesized frames. A low ALD score combined with a high code block error rate (CBLER) indicates that the $R(\cdot)$ is able to accurately estimate the missing landmarks. LPIPS is used to approximate the perceptual similarity between the ground truth and synthesized frames. Finally, SSIM approximates the ability of the synthesized frames to reproduce the ground truth frames.

B. Results

For the FOMM application task, we implement our framework as follows. A 512x512 resolution video of head and shoulders video conferencing is used as the information source I, with $E(\cdot)$ implemented using the FOMM motion estimation module [13], generating landmarks. The $P(\cdot)$ is implemented as a weighted average of landmark motion and landmark contribution to emotion detection, where the contribution is taken as the gradient of the cross-entropy loss w.r.t. to each landmark between detected emotion in the original frame, and the frame generated by those landmarks. PRT parameterization for the $\mathrm{UE_{TX}}$ to $\mathrm{BS_{UL}}$ is given in Table I, with nine tests. $R(\cdot)$ estimates the position of missing landmarks and is implemented as a simple fully connected network with two hidden layers of 80 neurons each, ReLU activation, and min-max normalized input and output. We train $R(\cdot)$ on a dataset of 1.6 million frames, with a uniform distribution of drop rates between 0 and 1. The $Syn(\cdot)$ is implemented with the FOMM generation module [13].

We evaluate the performance of FFH, SIDH, and CIDH on the FOMM task, against the objective function defined in Eq. 1 (W_ρ), and the average transmitted symbols per frame (μN), ie. how well, given the parameterization of K and C_j, the heuristics fulfill constrain Eq. 3. The results, averaged for CBLERs between 0 and 100%, are presented in Fig. I. CIDH consistently outperforms FFH and SIDH in W_ρ and μN, with CIDH achieving the highest W_ρ and the lowest μN across all tests. SIDH is only able to outperform FFH in W_ρ for single block allocation, and not at all in μN. By prioritizing high-importance symbols, SIDH is able to achieve a higher W_ρ than FFH. Further, because CIDH attends to symbol order, a high μN is achieved. Further finetuning of the CIDH threshold to roll over to unallocated symbols has the potential to improve the performance of CIDH, and is left for future work. In Fig. 2, the ALD, LPIPS, and SSIM scores for the FOMM model (for tests 1, 2, and 3 in Tab. I) are plotted. As shown, the plotted image metrics further validate the objective function as an approximation of QoE, with CIDH outperforming FFH and SIDH across all metrics for non-single block allocation. The superior performance of FFH in LPIPS and SSIM for single block allocation stems from the diversity in transmitted symbols over time, which benefits the error concealer. The initial worsening of LPIPS and SSIM betwen CBLERs of 100% and 90% is due to the error concealer defaulting to a base state under insufficient fresh semantics, reflected by the simultaenous improvement in ALD. In Fig. 3, the ALD, LPIPS, and SSIM for the tests parameterized in Tab. I are displayed for CIDH, juxtaposed with a baseline of 512x512 H.265 compressed video at 15kbps. Under the presence of block errors, H.265 compressed video is unable to decode, while PRT with CIDH is able to maintain perceptual quality. Compressed H.265 video is unable to outperform PRT on the ALD metric. This is due to the FOMM video synthesis operating on a source frame, yielding a clearer (even if incorrect) image, benefitting the landmark detection. Finally, we compare the bitrates of PRT with CIDH to H.265 compressed video at equivalent and higher bit rates in Fig. 4.

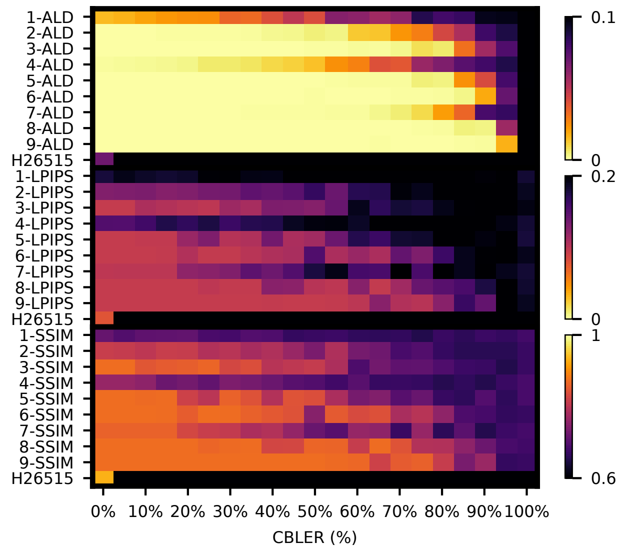

Fig. 3: FOMM ALD, LPIPS, and SSIM scores for the proposed CIDH strategy at varying Eb/N0 values and parameters (see Table I), compared to 256x256 H.264 compressed video at 30kbps.

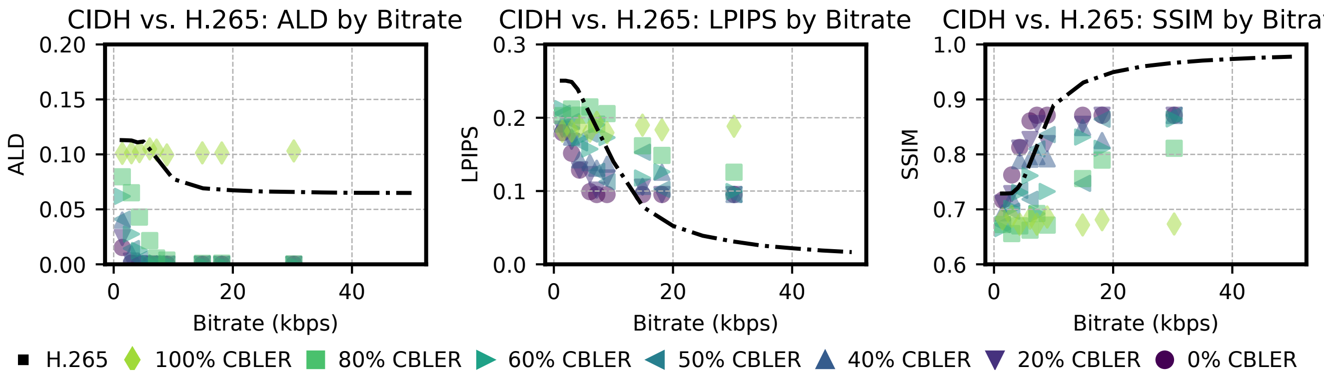

Fig. 4: ALD, **LPIPS** (lower is better), and **SSIM** (higher is better) scores for CIDH at varying bitrates vs. H.264 encoding.

While PRT is not able to outperform H.265 at every bitrate, significant gains in resilience to block errors are realized, as evidenced by the comparable performance of PRT at less than 50% CBLER with H.265.

VI. Conclusion and Outlook

We summarize the key contributions of this paper into three points. First, we propose a novel semantic-aware, importance-driven framework for semantic communication in NRLF environments. Second, we implement and evaluate our framework on a semantic video synthesis task, demonstrating significant improvements in robustness. Third, we show that our framework can revolutionize future wireless networks by ensuring high-quality service delivery in challenging transmission conditions at rates comparable to classical methods.

In this paper, we evaluated our framework with a video conferencing scenario, using the FOMM video synthesis model. While we demonstrate the ability of our framework to improve the robustness of semantic communication for the FOMM, we acknowledge that more rigourous, multi-modal evaluation is necessary to fully understand the potential of our framework. In future work, we plan to evaluate our framework on a wider range of video, speech, and natural language processing tasks, to demonstrate the flexibility and robustness of our framework in a variety of applications. Further, we believe that semantic importance-driven mechanisms are applicable to a wide range of wireless communication problems, such as QoS adjustments, routing decisions, or traffic management. Such mechanisms can easily be integrated into our framework.

References

[1] Emilio Calvanese Strinati and Sergio Barbarossa. 6g networks: Beyond shannon towards semantic and goal-oriented communications. *Computer Networks*, 190:107930, 2021.

[2] Cheng-Xiang Wang, Marco Di Renzo, Slawomir Stanczak, Sen Wang, and Erik G Larsson. Artificial intelligence enabled wireless networking for 5g and beyond: Recent advances and future challenges. *IEEE Wireless Communications*, 27(1):16–23, 2020.

[3] Qiao Lan, Dingzhu Wen, Zezhong Zhang, Qunsong Zeng, Xu Chen, Petar Popovski, and Kaibin Huang. What is semantic communication? a view on conveying meaning in the era of machine intelligence. *Journal of Communications and Information Networks*, 6(4):336–371, 2021.

[4] Peiwen Jiang, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. Wireless semantic communications for video conferencing. *IEEE Journal on Selected Areas in Communications*, 41(1):230–244, 2022.

[5] Goluck Konuko, Giuseppe Valenzise, and Stéphane Lathuilière. Ultra-low bitrate video conferencing using deep image animation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4210–4214. IEEE, 2021.

[6] Huiqiang Xie, Zhijin Qin, Geoffrey Ye Li, and Biing-Hwang Juang. Deep learning enabled semantic communication systems. *IEEE Transactions on Signal Processing*, 69:2663–2675, 2021.

[7] Yining Wang, Mingzhe Chen, Tao Luo, Walid Saad, Dusit Niyato, H Vincent Poor, and Shuguang Cui. Performance optimization for semantic communications: An attention-based reinforcement learning approach. *IEEE Journal on Selected Areas in Communications*, 40(9):2598–2613, 2022.

[8] Chuanhong Liu, Caili Guo, Yang Yang, and Nan Jiang. Adaptable semantic compression and resource allocation for task-oriented communications. *arXiv preprint arXiv:2204.08910*, 2022.

[9] Lei You, Qi Liao, Nikolaos Pappas, and Di Yuan. Resource optimization with flexible numerology and frame structure for heterogeneous services. *IEEE Communications Letters*, 22(12):2579–2582, 2018.

[10] Ali A Esswie and Klaus I Pedersen. Opportunistic spatial preemptive scheduling for urllc and embb coexistence in multi-user 5g networks. *Ieee Access*, 6:38451–38463, 2018.

[11] Renato Abreu, Preben Mogensen, and Klaus I Pedersen. Pre-scheduled resources for retransmissions in ultra-reliable and low latency communications. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–5. IEEE, 2017.

[12] Zhenzi Weng and Zhijin Qin. Semantic communication systems for speech transmission. *IEEE Journal on Selected Areas in Communications*, 39(8):2434–2444, 2021.

[13] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. First order motion model for image animation. *Advances in Neural Information Processing Systems*, 32:7137–7147, 2019.

[14] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *Proceedings of International Conference on Learning Representations, San Diego,*, 2015.

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[16] Kun Lu, Qingyang Zhou, Rongpeng Li, Zhifeng Zhao, Xianfu Chen, Jianjun Wu, and Honggang Zhang. Rethinking modern communication from semantic coding to semantic communication. *IEEE Wireless Communications*, 30(1):158–164, 2022.

[17] Joon Son Chung, Arsha Nagrani, and Andrew Zisserman. Voxceleb2: Deep speaker recognition. *arXiv preprint arXiv:1806.05622*, 2018.

[18] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[19] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.