Disruption-Resilient Real-Time Sensor Data Delivery via Neural Multiple Description Coding

Xinyue Hu, Qixin Zhang, Wei Ye, Eman Ramadan, Zhi-Li Zhang
University of Minnesota Twin Cities, Minneapolis, USA
{hu000007, zhan8548, ye000094, eman}@umn.edu zhzhang@cs.umn.edu

***Abstract*—In this paper we develop a novel *disruption-resilient* approach for *real-time*, *high-resolution* sensor data delivery over multiple wireless channels for military autonomous systems such as drones, autonomous vehicles and robots. We design two innovative *neural multiple description codecs* (neural MDCs) which compress and encode images into multiple *independently decodable* and *mutually refineable* streams. Our approach not only achieves high compression efficiency, but also enables the effective use of multiple diverse radio channels for real-time delivery of high-resolution sensor data while ensuring *disruption resiliency*. Using benchmark image/video sensor datasets as well as real-world 5G traces, we evaluate and demonstrate the efficacy of both neural MDC codecs for high-resolution sensor data streaming over multiple radio channels under various jamming scenarios.**

I. Introduction

Grounded and aerial autonomous systems such as unmanned ground vehicles (UGVs), unmanned aerial vehicles (UAVs or "drones") and robots are integral part of emerging and future warfare. Apart from allowing warfighters to engage in combat activities remotely, these autonomous systems are especially useful in providing surveillance and real-time situation awareness to assist effective command-and-control and decision-making. Due to processing and power constraints, autonomous systems will likely have limited AI (artificial intelligence) capabilities, e.g., by running "small AI models" that do not require large onboard memory and processing powers, that are primarily used for their autonomous operations. To equip them with advanced deep learning capabilities – especially *generative* AI, it will be necessary to connect the autonomous systems *wirelessly* to backend (edge or cloud) AI systems running state-of-the-art large models.

Streaming (real-time) sensor data – especially high-resolution camera images, video and 3D point data – faces several challenges. Wireless channel bandwidth is often limited and is known to suffer high variability. This is particularly the case when mobility is involved. For *real-time* situation awareness, ensuring timely delivery of sensor data with low latency is also critical. In military applications, intentional radio interference and signal jamming are common techniques used by adversaries to disrupt electronic communications. One effective approach to prevent channel disruption and jamming is to utilize multiple radio channels from different radio bands. For example, channel surfing and frequency hopping [1]–[3] are commonly used to evade jamming. These methods require tight synchronization and coordination of the sender and receiver. Further, if the current channel used for sensor data delivery is disrupted, retransmissions using a different channel will be needed, incurring longer delay which may render the sensor data obsolete. Another simple strategy is to replicate and transmit the same sensor data simultaneously using multiple channels. This not only makes jamming in general more difficult[1], but also reduce the need for retransmissions. However, this strategy wastes valuable radio capacity, and may not be suited for high-resolution sensor data delivery when the bandwidth of individual channels is insufficient to transmit the sensor data in a timely manner.

In this paper we develop a novel *disruption-resilient* approach for *real-time*, *high-resolution* sensor data delivery over multiple wireless channels for military autonomous systems such as UAVs and UGVs. By leveraging recent advances in neural image codecs, we advocate the use of *neural multiple description coding* (neural MDC) to compresses and encodes images into multiple *independently decodable* and *mutually refineable* streams for disruption resiliency. In particular, we design two neural MDC codecs, Pixel MDC and TokenMDC, for effective high-resolution image compression and streaming over multiple radio channels. Both neural codecs not only achieve high compression efficiency, but also enable the effective use of multiple diverse radio channels for real-time delivery of high-resolution image sensor data while ensuring *disruption resiliency*. Using visual quality, object detection accuracy and stall ratios as key metrics and nuScenes and VisDrone benchmark sensor datasets as well as real-world 5G channel throughput traces, we evaluate and demonstrate the efficacy of both neural MDC codecs for high-resolution sensor data streaming over multiple radio channels subject to various jamming scenarios. The key results are summarized below:

- PixelMDC and TokenMDC both retain high visual quality and detection accuracy when only one or two channels are disrupted, and maintain good graceful performance degradation when more channels are disrupted.
- TokenMDC provides more favorable compression-detection trade-offs, saving 29.5% bitrate compared to

[1]Constantly and simultaneously jamming multiple channels operating on diverse frequency bands not only require high power and more advanced equipment/technologies; the high power used also likely exposes the location of a jammer, risking detection and elimination [4]–[7].

979-8-3315-0292-8/25/$31.00 ©2025 IEEE

JPEG and 67.3% compared to PixelMDC for similar detection accuracy.
- TokenMDC outperforms PixelMDC in terms of loss resiliency, attaining 27% higher MS-SSIM of reconstruncted images with losses.
- Both PixelMDC and TokenMDC achieve excellent real-time streaming performance, with stall ratios below 0.81% and 0.11%,respectively, across all jamming scenarios.
- Overall, TokenMDC achieves the best trade-offs between low latency, visual quality, detection accuracy and disruption resiliency.

II. Background & Related Work

We provide a brief background on jamming and neural codes and contrast our approach with related work.

A. Jamming Attacks, Detection and Mitigation Strategies

Jamming is a classic technique in electronic warfare [6], [7]. There is a vast research literature on jamming attacks, their detection and prevention strategies (see, e.g., [1]–[3], [8] and the references therein). Depending on the perspectives used, jamming attacks may be broadly classified into i) passive vs. active jammers (the latter includes constant, deceiptive and random jammers) [1]; ii) general vs. function-specific jammers [3] where the latter targets specific functions such as time synchrononization, control channels, etc.; and iii) spot jamming, barrage jamming, sweep jamming and digital frequency radio memory (DFRM) [4], [5], where the first two target a specific channel either randomly or constantly, the third aims to jam multiple frequencies in quick succession, although not all at the same time; whereas the last one is a repeater technique typically used for jamming radar signals – it alters and re-transmits received radar energy to confuse a radar. The recent survey paper [3] provides in-depth discussion of various jamming attacks, their detection and mitigation methods for various types of wireless networks. In military applications, a jammer needs to weigh effectiveness of jamming vs. the possibility of detection (thereby risking being destroyed). For example, sweep jamming spreads power across multiple frequencies, making it comparatively less powerful at a single frequency, whereas barrage jamming has a high probability of being detected [4]–[7].

Most anti-jamming studies either employ PHY/MAC and resource allocation techniques or focus on mitigating function-specific jamming. Some recent studies leverage AI methods for jamming detection and mitigation (see, e.g., [8]), whereas others target specific use cases (e.g., drones [9], [10] or autonomous vehicles [11]). In this work, we advocate a novel general approach based on innovative neural codecs that employs multiple diverse channels (instead of channel surfing/hopping) to simultaneously achieve high bandwidth, low latency, application performance and disruption resiliency.

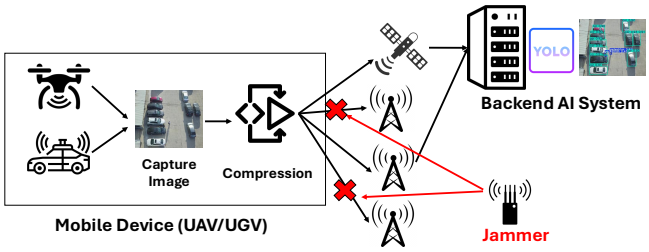


Fig. 1: System overview of sensor data delivery to a backend AI system over diverse radio channels under jamming, enabling advanced deep vision tasks: an example where two uplink channels are disrupted by a jammer.

B. Neural Image/Video Codecs

Image/video compression is a widely studied and mature field. Multiple Description Coding (MDC) using classical signal processing techniques have been proposed decades ago [12]; however, due to inefficiency and high overheads, it has never become practical. With rapid advances in AI, deep learning-based *neural* image and video codec designs have seen a flurry of activities in recent years (see, e.g., [13], [14]). In contrast to these studies, our work revisits the idea of MDC for its anti-jamming properties. Inspired by our earlier work on neural MDC codecs for video streaming over 5G networks [15], this work develops two novel neural MDC image codecs for disruption-resilient, real-time delivery of high-resolution images using diverse radio channels, with the goal to support military autonomous systems for real-time situation awareness, object detection, tracking and other mission-critical tasks.

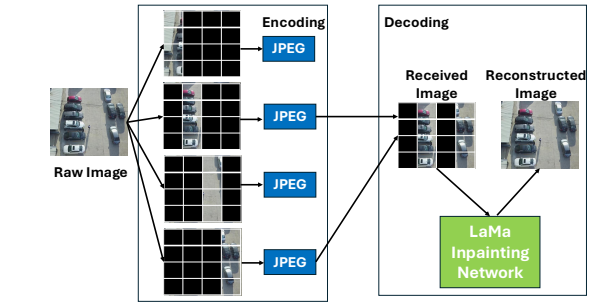
III. Methodology

A. System Overview

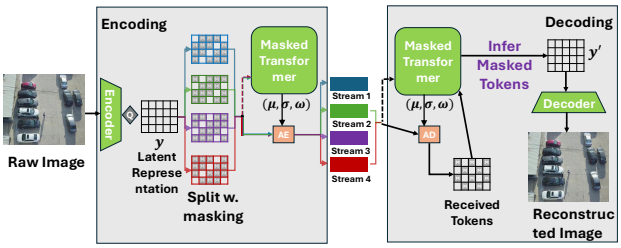
Fig. 1 illustrates the system overview of our framework, which utilizes diverse radio channels to transmit sensor data under jamming attacks, enabling advanced deep vision tasks on resource-constrained UAVs and UGVs via a backend AI system. First, UAVs or UGVs capture raw images via onboard cameras. Due to their limited computational and battery resources, these devices offload advanced vision tasks (*e.g.,* object detection) to a more powerful backend AI system over radio uplinks. However, the uplink bandwidth of a single radio channel is ofen limited and particularly vulnerable to jamming attacks. To ensure sufficient data throughput, the system employs multipath streaming over multiple radio channels, providing both increased bandwidth and resilience to jamming. Crucially, we employ a loss-resilient compression technique that reduces image size to accommodate limited bandwidth and tolerates partial data loss due to jamming without requiring retransmission, thereby enabling real-time streaming, object detection, tracking and other mission-critical tasks.

B. Loss Resilient Compression

MDC is a loss-resilient compression method that encodes an image into multiple *independently decodable* and *mu-*



(a) PixelMDC



(b) TokenMDC

Fig. 2: Overview of PixelMDC and TokenMDC, exemplifying the generation of four descriptions and successful decoding from two received descriptions.

tually refinable streams. The key design principles of an MDC image codec are: 1) determining the type of source information (e.g., spatial regions, semantic features) from the original image to distribute across multiple correlated descriptions with intentional redundancy; 2) leveraging this redundancy to recover lost descriptions from the received ones.

Inspired by recent advances in generative AI for images and vision, we design two MDC image codecs by splitting different types of source information, pixel blocks and latent tokens, into multiple descriptions. We leverage two distinct generative models, the LaMa image inpainting network [16] and a Masked Transformer [13], to infer lost descriptions. Fig. 2 illustrates the frameworks of the two codecs: PixelMDC and TokenMDC.

*1) **PixelMDC**:* We extend traditional spatial MDC codecs, which interleavingly distribute pixel blocks of an image into multiple descriptions and encode each description independently using JPEG. Unlike prior methods, we incorporate LaMa, a state-of-the-art image inpainting network, to reconstruct missing blocks and enhance loss resilience.

*2) **TokenMDC**:* We design TokenMDC based on recent Masked Transformer-based neural codecs [13], [15]. It consists of three key components: an AutoEncoder, a source information splitting and merging module, and a Masked Transformer. Given a raw input image, TokenMDC first uses the AutoEncoder [17] to transform the image from the pixel domain into a quantized latent representation $\mathbf{y}$. Because the latent tokens are spatially correlated after the AutoEncoder transform, TokenMDC splits the tokens into multiple descriptions by interleavingly masking portions of $\mathbf{y}$ with a special learnable mask token. This process creates multiple masked latent representations, whose combination equals the original. According to Shannon's source coding theorem, the more accurately the distributions of the latent tokens are estimated, the fewer bits are required to transmit them. Motivated by the superior performance of Masked Transformers in image generation, TokenMDC employs a Masked Transformer to estimate the token distributions for each description. These estimated distributions, combined with arithmetic coding, are then used to compress each description into a bitstream to transmit over cellular networks.

At the receiver side, each MDC stream is independently decodable since each one is independently entropy encoded. When some MDC streams are lost, their missing tokens are filled with the mask token and inferred by the Masked Transformer using the received tokens and the estimated token distributions. Any combination of received streams contributes to the reconstructed latent representation's accuracy and improves the decoded image's quality. Further details can be found in the technical paper [15].

C. Multipath Streaming against Channel Disruption

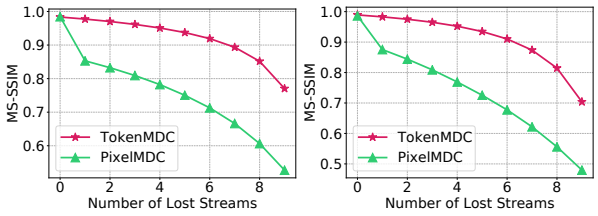
Instead of transmitting a single MDC stream across multiple uplink channels, we assign each MDC stream to a separate channel. This design ensures that the delivery of each stream depends solely on the condition of its assigned channel, making it resilient to disruptions on other channels. We estimate the available bandwidth using the most recent sending rate and compress images to match this estimated bitrate. When jamming or interference disrupts some channels, or when available bandwidth is overestimated, TokenMDC and PixelMDC reconstruct the image as long as at least one MDC stream is received, without waiting for delayed streams or retransmitting lost streams. This enables real-time, robust, and bandwidth-adaptive streaming. In contrast, JPEG-based streaming requires full bitstream reception for successful decoding, meaning any lost or delayed data must be retransmitted or waited on, leading to significant delays under fluctuating network condition or jamming/interference.

D. Object Detection

We uses object detection as a representative case study of advanced deep vision tasks. The backend (edge or cloud) server runs YOLO [18], a state-of-the-art deep learning model for object detection. YOLO is a single-stage detector that predicts bounding boxes and class probabilities directly from the input image in a single forward pass, making it well suited for time-sensitive applications such as autonomous navigation and remote monitoring. It remains robust to degraded inputs by preserving structural features such as overall shape and spatial patterns, even when fine details are lost.

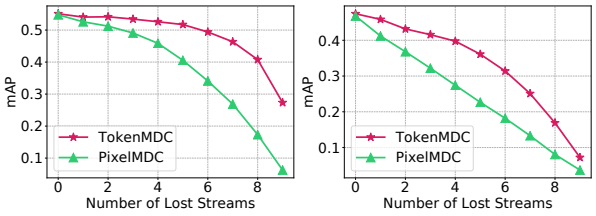
IV. Experimental Results

This section evaluates: (1) the rate-distortion and loss resilience performance of neural multiple description codecs



(a) nuScenes (b) VisDrone

Fig. 3: Impact of lost streams on visual quality.



(a) nuScenes (b) VisDrone

Fig. 4: Impact of lost streams on object detection accuracy.

and their impact on object detection; and (2) the streaming and object detection performance over 5G under jamming.

A. Experiment Setup

Multimedia Dataset: We conduct evaluations on the nuScenes [19] and VisDrone [20] datasets. nuScenes provides high-resolution 360-degree camera images captured from autonomous vehicles in urban settings, with annotations of vehicles, pedestrians, and other road users. VisDrone contains aerial images taken by drones in urban and suburban areas, containing small objects and complex backgrounds. These datasets represent complementary terrestrial and aerial use cases, allowing for a comprehensive evaluation of our streaming and detection pipeline.

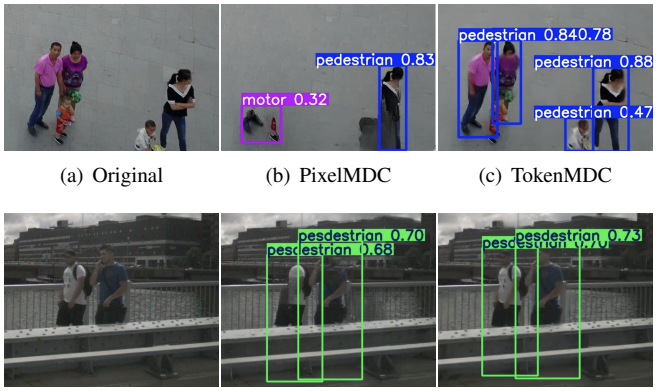
Network Trace: We use Xcal to collect the unlink network traces from major U.S. 5G operators concurrently. We randomly select 4 channel traces from the collected traces to emulate four-path multipath scenarios, resulting in total 20 test cases.

Evaluation Metrics: We use bits per pixel (BPP) to measure the image size after compression and Multi-Scale Structural Similarity Index Measure (MS-SSIM) to measure the visual quality of an image. Streaming performance is evaluated based on the trade-off between visual quality and real-timeness, measured by stall time (*i.e.,* the transmission delay beyond the expected decoding time based on the streaming frame rate). Based on decoded images, mean Average Precision (mAP) is used to quantify object detection accuracy.

Testbed: We implemented TokenMDC based on M2T [13], and PixelMDC using FFmpeg JPEG and LaMa [16]. We use these two MDC codecs to compress images into 10 independent streams. We conducted multipath streaming in a controlled environment by replaying channel bandwidth traces using Mahimahi. During streaming, a simulated jammer randomly targets 1 to 3 channels with concentrated jamming signals to disrupt data transmission. The mobile devices (*i.e.,* vehicle and drones) send images at the frame rate of 12. On the edge server, two YOLOv8s [18] models, fine-tuned on the nuScenes and VisDrone datasets respectively, are used to detect objects from the images decoded from the received bitstreams. We measure transmission delay by accounting for the network delay, neural codec runtime, LaMa inpainting runtime, and retransmission overhead.

B. Compression Performance & Impact on Object Detection

1) ***Loss Resilience****:* MDC coding schemes are inherently loss resilient, as the image can still be decoded—albeit



(a) Original (b) PixelMDC (c) TokenMDC

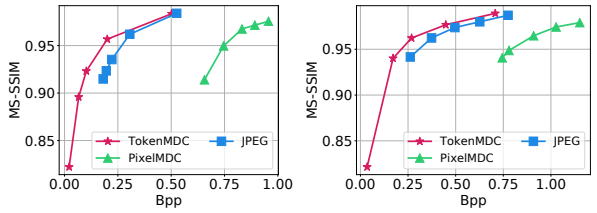
(d) Original (e) PixelMDC (f) TokenMDC

Fig. 5: Image reconstruction and object detection examples when 50% of the streams are lost.

with reduced quality—as long as one or more streams are received. We evaluate the loss resilience of our two MDC codecs by measuring visual quality and object detection accuracy under various stream loss ratios (see Fig. 3 and Fig. 4). TokenMDC outperforms PixelMDC, achieving 27% higher MS-SSIM, indicating more effective inference of lost tokens from received ones compared to pixels. This improvement translates to a 72% increase in object detection accuracy. When comparing performance across the two datasets, we observe that nuScenes images exhibit greater loss resilience than VisDrone images, as the small, aerially captured objects in VisDrone are more sensitive to loss and harder to reconstruct accurately.

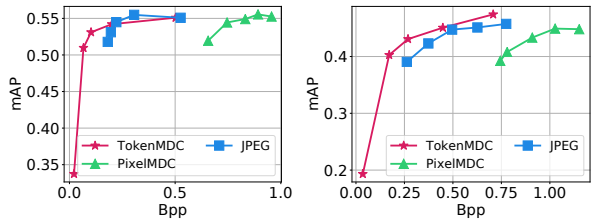
Fig. 5 shows the reconstructed images and their object detection results when 50% of the MDC streams are lost. TokenMDC produces reconstructions that are visually closer to the original images, preserving object contours even though fine details are blurred. These structural features help YOLO detect objects more accurately and in greater numbers, despite the data loss. In contrast, PixelMDC performs well when large objects are present (e.g., Fig. 5 (e)), but struggles to recover small objects if their corresponding pixel blocks are lost. This is because pixel blocks carry only limited local spatial information, whereas tokens in TokenMDC encode richer semantic context from the entire image. When no semantically similar blocks are available (e.g., people in Fig. Fig. 5 (b)), even advanced inpainting models like LaMa cannot reliably reconstruct them.

2) ***Rate-Distortion****:* Fig. 6 compares the compression efficiency of our two MDC codecs with JPEG. TokenMDC



(a) nuScenes (b) VisDrone

Fig. 6: Impact of bitrate on visual quality.



(a) nuScenes (b) VisDrone

Fig. 7: Impact of bitrate on object detection accuracy.

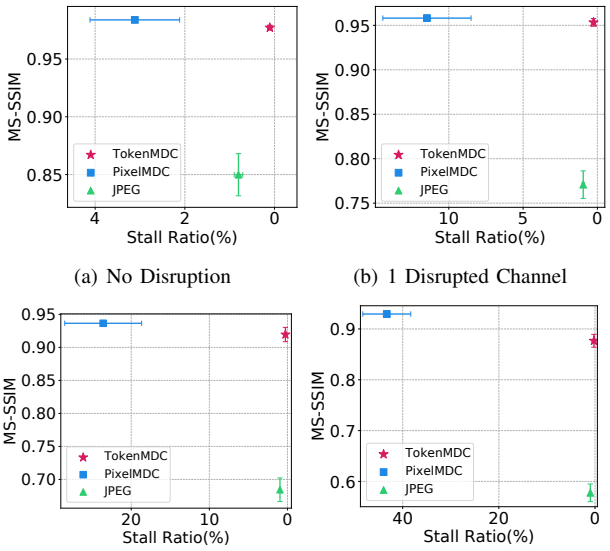
achieves 34.3% higher compression efficiency than JPEG and 70.1% higher than PixelMDC. Additionally, TokenMDC can compress images to a bitrate 87.2% lower than JPEG's lowest quality setting, which indicates that TokenMDC can significantly reduce network bandwidth consumption. This high compression efficiency improves adaptability to low-bandwidth conditions caused by poor channels or jamming, enabling real-time streaming, and also reduces transmission costs, making it more economical for usage-based or bandwidth-constrained networks. The superior rate-distortion performance of TokenMDC stems from its use of masked Transformers, which capture richer spatial relationships among tokens and enable more accurate token distribution prediction for improved entropy coding. In contrast, PixelMDC's reduced spatial redundancy and reliance solely on JPEG limit its ability to exploit spatial correlations among pixels, leading to much lower compression efficiency.

3) ***Rate-Detection****:* Fig. 7 presents object detection accuracy for images compressed at various bitrates. It illustrates the trade-off between compression and detection performance by showing how bitrate translates to mAP. TokenMDC achieves more favorable compression-detection trade-offs, saving 29.5% bitrate compared to JPEG and 67.3% compared to PixelMDC for similar detection accuracy. Detection accuracy on the VisDrone dataset is lower than on nuScenes and degrades more rapidly with decreasing bitrate. This is because small drone-captured objects are harder to detect and more sensitive to compression.

C. Streaming & Detection Performance Against Disruption

This section evaluates the multipath streaming and object detection performance under jamming attacks and interference that randomly disrupt uplink channels.

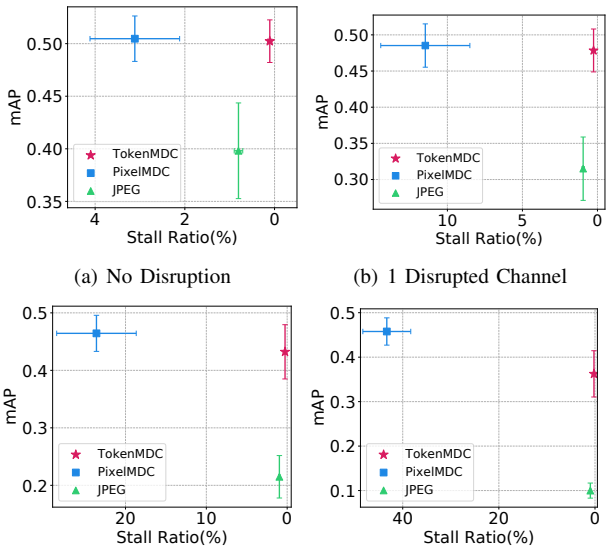
1) ***Streaming Peformance****:* Fig. 8 shows that streaming nuScenes and VisDrone images to the edge server using PixelMDC and TokenMDC achieves real-time performance,



(a) No Disruption (b) 1 Disrupted Channel

(c) 2 Disrupted Channels (d) 3 Disrupted Channels

Fig. 8: Streaming performance across four channels under different jamming scenarios. Streaming with MDC achieves real-timeness. Error bars show 95% confidence intervals.



(a) No Disruption (b) 1 Disrupted Channel

(c) 2 Disrupted Channels (d) 3 Disrupted Channels

Fig. 9: Object detection performance across four channels under different jamming attacks. Streaming with TokenMDC achieves the best trade-offs between real-timeness and detection accuracy. Error bars show 95% confidence intervals.

with stall ratios below 0.81% and 0.11%, respectively, across all disruption scenarios. This is because both MDC codecs can reconstruct an image as long as at least one stream is received, even if others are lost due to channel disruption. In contrast, streaming with JPEG experiences significant delays, particularly as the number of disrupted channels increases. Since JPEG requires complete bitstream reception for successful decoding, lost data must be retransmitted over the remaining undisrupted channels, which adds delay. Disrupted channels reduce the aggregate network throughput

below the planned upload rate, further slowing transmission. These results also highlight JPEG's limited ability to adapt to dynamic network conditions under jamming.

While maintaining real-time performance, PixelMDC and TokenMDC codecs sacrifice visual quality under stream loss. However, TokenMDC achieves 14.97% to 51.79% higher visual quality than PixelMDC, owing to its superior rate-distortion and loss resilience capabilities, as shown in the previous two subsections.

*2) **Detection Peformance**:* Fig. 9 demonstrates that TokenMDC offers the best trade-off between detection accuracy and real-time performance. In the absence of jamming or interference, TokenMDC matches JPEG in detection accuracy but achieves significantly lower latency (0.1% vs. 3.16%), due to its superior compression efficiency that allows higher-quality images at lower bitrates, despite its higher runtime. As channel disruptions increase, detection accuracy drops across all codecs. While JPEG's accuracy degrades slightly, its latency rises sharply, up to 45.22%, due to retransmission delays. In contrast, TokenMDC and PixelMDC maintain real-time performance under disruption, as they eliminate the need for retransmission. TokenMDC is notably more robust than PixelMDC. The accuracy of TokenMDC drops only 4.76% with one of four channels disrupted (∼25% data loss), 13.96% with two channels disrupted, and 27.85% with three. This accuracy decline is much slower than the rate of data loss, confirming that TokenMDC ensures real-time detection with graceful accuracy degradation under adverse conditions.

V. Conclusion

We advocate a novel *disruption-resilient* approach for *real-time*, *high-resolution* sensor data delivery over multiple wireless channels for military autonomous systems. We have developed two innovative *neural multiple description coding* (neural MDC) codecs – PixelMDC and TokenMDC – both of which compress and encode images into multiple *independently decodable* and *mutually refineable* streams. Using benchmark sensor datasets and real-world 5G traces, we have evaluated the performance of Pixel and Token MDC codecs for high-resolution sensor data streaming over multiple radio channels under various jamming scenarios. The evaluation results demonstrate the efficacy of our approach. In particular, TokenMDC achieves the best trade-offs between low latency, visual quality, detection accuracy and disruption resiliency.

VI. Acknowledgment

This research is supported in part by the National Science Foundation (NSF) under grants number 2106771, 2128489, 2212318, 2220286, 2220292, 2321531 and 2436333, as well as an InterDigital gift.

References

[1] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*, 2005, pp. 46–57.

[2] A. Siemens and M. van Hecke, "Jamming: A simple introduction," *Physica A-statistical Mechanics and Its Applications*, vol. 389, pp. 4255–4264, 2010.

[3] H. Pirayesh and H. Zeng, "Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey," *IEEE communications surveys & tutorials*, vol. 24, no. 2, pp. 767–809, 2022.

[4] C. Insight, "How jamming attacks work: A breakdown of the three types," 2021. [Online]. Available: https://cyberinsight.com/how-jamming-attacks-work-a-breakdown-of-the-three-types/

[5] J. Engineering, "An introduction to jammers and jamming techniques," 2020. [Online]. Available: https://jemengineering.com/blog-an-introduction-to-jammers/

[6] C.-H. Cheng and J. Tsui, *An introduction to electronic warfare; from the first jamming to machine learning techniques*. CRC Press, 2022.

[7] L. Boudreaux and U. Army, *COMMUNICATIONS JAMMING HANDBOOK*. Independently published, 2021.

[8] P. Lohan, B. Kantarci, M. A. Ferrag, N. Tihanyi, and Y. Shi, "From 5g to 6g networks, a survey on ai-based jamming and interference detection and mitigation," *IEEE Open Journal of the Communications Society*, 2024.

[9] L. Ye, J. Zhang, H. Chen, Z. Lin, J. Li, Z. Lv, and L. Xiao, "Learning-based edge-assisted uav object detection against jamming for extended reality," in *2024 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2024, pp. 1287–1292.

[10] Z. Lv, L. Xiao, Y. Du, G. Niu, C. Xing, and W. Xu, "Multi-agent reinforcement learning based uav swarm communications against jamming," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9063–9075, 2023.

[11] Q. Zhang, S. Sleder, X. Hu, F. Bilal, W. Ye, and Z.-L. Zhang, "Impact of data compression on downstream ai tasks: A study using teleoperated driving over 5g," in *2024 IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR)*. IEEE, 2024, pp. 25–30.

[12] M. Kazemi, S. Shirmohammadi, and K. H. Sadeghi, "A review of multiple description coding techniques for error-resilient video delivery," *Multimedia Systems*, vol. 20, pp. 283–309, 2014.

[13] F. Mentzer, E. Agustson, and M. Tschannen, "M2t: Masking transformers twice for faster decoding," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 5340–5349.

[14] J. Li, B. Li, and Y. Lu, "Neural video compression with diverse contexts," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Vancouver, BC, Canada: IEEE, 2023, pp. 22 616–22 626.

[15] X. Hu, W. Ye, J. Tang, E. Ramadan, and Z.-L. Zhang, "Robust multiple description neural video codec with masked transformer for dynamic and noisy networks," *arXiv preprint arXiv:2412.07922*, 2024.

[16] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky, "Resolution-robust large mask inpainting with fourier convolutions," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 2149–2159.

[17] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, "Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5718–5727.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[19] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[20] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 11, pp. 7380–7399, 2021.