2025 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)

2025 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) | 979-8-3315-3553-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/PerComWorkshops65533.2025.00074

TAVIC-DAS: Task and Channel-Aware Variable-Rate Image Compression for Distributed Autonomous System

Gaurav Shinde[†], Anuradha Ravi[†], Emon Dey[†], Jared Lewis[†], Nirmalya Roy[†]
[†]*University of Maryland, Baltimore County, USA*
[†]{gshinde1, anuradha, edey1, jlewis9, nroy}@umbc.edu

Abstract*—In network-constrained environments, distributed multi-agent systems—such as UGVs and UAVs—must communicate effectively to support computationally demanding scene perception tasks like semantic and instance segmentation. These tasks are challenging because they require high accuracy even when using low-quality images, and the network limitations restrict the amount of data that can be transmitted between agents. To overcome the above challenges, we propose *TAVIC-DAS* to perform a task and channel-aware variable-rate image compression to enable distributed task execution and minimize communication latency by transmitting compressed images. *TAVIC-DAS* proposes a novel image compression and decompression framework (distributed across agents) that integrates channel parameters such as RSSI and data rate into a task-specific "semantic segmentation" DNN to generate masks representing the object of interest in the scene (ROI maps) by determining a high pixel density needed to represent objects of interest and low density to represents surrounding pixels within an image. Additionally, to accommodate agents with limited computational resources, *TAVIC-DAS* incorporates resource-aware model quantization. We evaluated *TAVIC-DAS* on platforms such as ROSMaster X3 and Jetson Xavier, which communicated using a low-frequency proprietary Doodle radio operating at 915 MHz. The experimental results show that *TAVIC-DAS* achieves approximately 7.62% higher PSNR and is about 6.39% more resource efficient compared to state-of-the-art techniques.

***Index Terms*—Task-Aware Image Compression, Edge Computing, Scene Understanding, Pervasive Computing**

I. Introduction

Autonomous agents such as Unmanned Aerial Vehicles (UAVs), and Unmanned Ground Vehicles (UGVs) collaborate on critical tasks like path planning and scene understanding. However, these devices are inherently resource-constrained, with limited computational power, memory, and energy, and communicate in network-constrained environments. The agents transmit images of the perceived scenes among each other to facilitate efficient information exchange and enhance the accuracy of contextual interpretation. Thus, image compression is pivotal in (a) reducing the data size communicated over the network to minimize transmission latency [1] and (b) lowering the computational requirements to process the images for various tasks [2]. While lossless image compression is ideal for preserving information, lossy compression significantly reduces file size by up to 90%, making it ideal for network-constrained environments. Traditional compression techniques, including Huffman coding, Slepian-Wolf coding [3] along with deep learning-based approaches like Generative Adversarial Networks (GANs) [4], have been widely employed for image compression. These methods are the foundation for more advanced techniques like variable-rate image compression, which enables dynamic adjustment of compression rates, allowing a single model to operate efficiently across varying image bitrates. To enable variable-rate compression for various image tasks, Pu et.al [5] proposed conditional class entropy-based metrics that are incorporated into an encoder to create compressed images, which are then decompressed for downstream tasks. Current state-of-the-art image compression techniques often assume stable and high-bandwidth network conditions, which are not true for real-world contested environments involving autonomous agents. Consequently, these approaches are unsuited for network-constrained environments that depend on low-frequency and long-range communication technologies, such as LoRa [6] or proprietary systems like Doodle Radios. This motivates researchers towards joint task and channel-aware image compression and transmission between autonomous agents [7]. Given the resource constraints of the agents, the processes of image compression, transmission, and task execution must prioritize power efficiency and minimize latency. While image compression performed using transformers ([8], [9]) may have high accuracy, they are not resource-efficient. State-of-the-art joint task and channel-aware models [7] do not propose unified models for different compression rates. To overcome the above challenges, we propose *TAVIC-DAS* , a *joint task and channel-aware variable-rate image compression and transmission framework for resource-constrained autonomous agents*.

Keeping the task of "instance segmentation" in focus, *TAVIC-DAS* distributes tasks among autonomous agents, leveraging shared image data to enhance collective scene perception. To ensure resource efficiency, *TAVIC-DAS* begins with task-specific downsampling of the captured image. It then utilizes a device-specific quantized "semantic segmentation" model at the sender to extract channel-aware region-of-interest (ROI) maps from the downsampled image. The pixel density in the ROI maps, determined by the α parameter, adapts to network conditions such as Received Signal Strength Indicator (RSSI) and bitrate. Under poor network conditions, the maps are sparsely populated (lower α), while good network conditions result in densely populated maps (higher α). Using these ROI maps, *TAVIC-DAS* compresses images in an 80:20 ratio, preserving 80% of pixels representing objects of interest and only preserving 20% of surrounding pixels, minimizing the loss of task-relevant details. The compressed image, along with the α parameter, is transmitted to the receiver, where it is reconstructed for instance segmentation. By pre-determining the regions of interest at the sender, *TAVIC-DAS* significantly reduces the computational load for performing

2766-8576/25/$31.00 ©2025 IEEE
DOI 10.1109/PerComWorkshops65533.2025.00074

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on July 03,2025 at 16:27:53 UTC from IEEE Xplore. Restrictions apply.

instance segmentation tasks. Key Contributions of *TAVIC-DAS* :

- **Joint Task and Channel Aware Image Compression.** Integrates network parameters for dynamic image compression based on network conditions. *TAVIC-DAS* achieves 69% reduction of image size with downsampling and 82.5% reduction of image size with channel-guided ROI map generation.
- **Resource Efficient Framework** *TAVIC-DAS* quantizes the task-specific model to optimize performance while minimizing memory and computational overhead. We show that names achieve 11.75% improvement in computation and 25.6% in communication power efficiency compared to state-of-the-art techniques. The distributed task execution (semantic-segmentation as upstream and instance-segmentation as downstream) effectively utilizes the resources across the agents.
- **Novel decompression framework** of *TAVIC-DAS* helps reconstruct the images that improves the IOU of instance segmentation task by $\approx$ 4% when compared to using compressed images.

II. Related work

In this section, we provide a concise overview of existing research methods for task and channel-aware image compression.

A. Task-aware image compression

Task-aware networks are extensively utilized in image compression and super-resolution, leveraging trained feature extractors to enhance the accuracy of downstream task inference. [10] proposed a variable-rate image compression framework guided by a quality map, allowing classification-aware compression by optimizing the map during encoding without retraining the model. Their method approximates the map on the decoder side using the latent representation. In contrast, *TAVIC-DAS* minimizes the payload by transmitting only the quality value(α), which serves as the sole conditioning factor for reconstruction. [11] used task-specific optimization of quantization tables through a differentiable loss function and demonstrated improved performance across multiple semantic and conventional compression tasks. Building on task-aware methods, [12] explored task-specific downscaling for super-resolution and colorization, emphasizing optimized reconstruction while maintaining compression efficiency. [13] proposed a recognition-aware learned compression method, which optimizes a rate-distortion loss alongside a task-specific loss, jointly learning compression and recognition networks. Task-aware approaches have also been explored in video compression, with [14] dynamically optimizing macro-block quantization parameters using reinforcement learning for tasks like car detection and ROI encoding and [15] proposing an encoder-controlled deep video compression framework that adapts pre-trained decoders for diverse machine vision tasks.

B. Channel-aware image compression

Channel-aware image compression considers channel parameters such as bandwidth, latency, and signal quality to enable differential compression that dynamically optimizes bit allocation based on the transmission environment. [7] considered channel gain, transmission rate, and latency to select semantic blocks for transmission, whereas *TAVIC-DAS* transmits the entire image using adaptive compression, dynamically adjusting quality based on bandwidth and latency to ensure task performance. This preserves the neighboring pixel information, which may be significant not just for downstream instance segmentation tasks but also if the receiver decides to execute other tasks, such as gesture recognition. [16] proposed a data-driven underwater image transmission system incorporating channel-aware training using an autoencoder to compress and encode images by leveraging underwater acoustic channel properties such as low bandwidth and variable path loss. Some work has explored channel parameters in compression sensing, which reconstructs signals from sparse samples, but image compression, which encodes full signals efficiently, has seen little focus on channel-aware methods.

III. Background and Motivation

Inspired by [10], *TAVIC-DAS* generates ROI maps for different rates of α depending on the channel parameters. However, the existing work involves transmitting a latent representation of the quality map. As we show, the higher alpha values that regulate the compression level of different regions significantly increase transmission latency and payload. Figure 1 illustrates that an increase in the alpha value corresponds to larger compressed file sizes, which in turn leads to higher latency. *It is important to note that the observed trend for file size and latency remains consistent as alpha increases, regardless of whether the image size is KB or MB. This consistency is attributed to higher alpha values prioritizing the preservation of finer details, thus necessitating a greater amount of data for encoding and transmission.* Furthermore, the increase in payload also increases the likelihood of retransmissions in TCP due to network congestion. This motivates us to ① transmit only the alpha value for image reconstruction at the receiver and ② incorporate latency and retransmissions to determine the optimal alpha value.

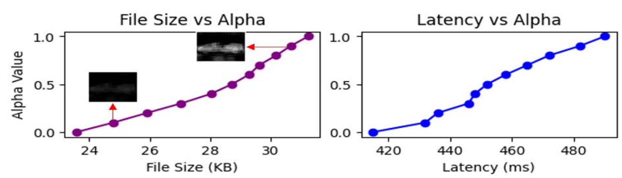


Fig. 1: File size increases as alpha increases. ROI maps corresponding to α = 0.2 and α = 0.9 are shown (best when zoomed in) (left). The transmission latency also increases with alpha (right).

IV. Methodology

This section outlines the design details of *TAVIC-DAS* .

System Overview

To enable a resource-efficient framework, *TAVIC-DAS* begins by downsampling the input image, enhancing computational efficiency for selecting objects of interest. Simultaneously, *TAVIC-DAS* evaluates network conditions using RSSI values and bitrate to derive an α value, which determines the pixel density to preserve for the downstream task. The downsampled image and α value are then used to generate a channel-guided, task-aware ROI map for the semantic segmentation task. The ROI map identifies objects of interest in the image and specifies the required pixel density. Leveraging this map, the compression module efficiently compresses the downsampled image and transmits it to the receiver along with the α value. At the receiver, the image is reconstructed, optimizing the Intersection over Union (IoU) for the instance segmentation

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on July 03,2025 at 16:27:53 UTC from IEEE Xplore. Restrictions apply.

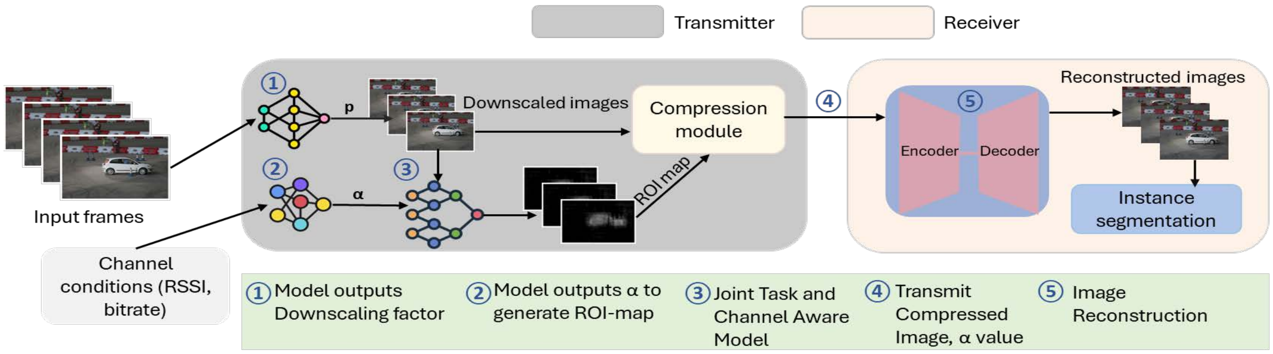


Fig. 2: System architecture of *TAVIC-DAS*

task. This pipeline (illustrated in Figure 2) enables task-aware compression, efficient transmission, distributed task execution, and precise instance segmentation at the receiver.

1) Training Phase

The training phase is conducted at both the transmitter and receiver sides, with the details outlined in this section.

Downscaling network. To perform the downscaling process, each input image $X \in \mathbb{R}^{m\times n}$ is resized using downscaling factors p ranging from 1 to 4, with a step size of 0.1. For each factor p, the downscaled dimensions are computed as $m' = \frac{m}{p}$ and $n' = \frac{n}{p}$, resulting in a resized image $X_{\text{downscaled}} \in \mathbb{R}^{m'\times n'}$. For every downscaled image, two losses are calculated: the rate loss L_R, defined as the estimated bits per pixel (BPP), and the semantic segmentation loss at the transmitter L_S, computed as the Intersection-over-Union (IoU) loss using the DeepLabV3-ResNet50 model [17]. The overall loss function is formulated as:

$$\mathcal{L} = L_R + \lambda L_S \tag{1}$$

where λ is a balancing factor that trades off between compression efficiency and segmentation accuracy. The MobileNetV2 model [18], chosen for its low computational cost, is trained to minimize this loss function. Given an input image, the trained model predicts the optimal downscaling factor p that balances compression efficiency and segmentation quality. This approach is inspired by [19], but we adapt it specifically for the semantic segmentation task.

Task-aware network. We take inspiration from [10] but adapt it specifically for semantic segmentation at the transmitter. To identify regions of interest for the downstream task at the receiver, we utilize DeepLabV3_ResNet50 in combination with Gradient-weighted Class Activation Mapping (Grad-CAM) [20]. Grad-CAM generates a saliency map $G \in \mathbb{R}^{m'\times n'}$, highlighting the most relevant areas of interest for the downstream task, with each value $G_{i,j}$ indicating the importance of pixel (i,j). The ROI map Q is then defined as:

$$Q_{i,j} = \alpha \cdot G_{i,j}, \quad \alpha \in [0,1] \tag{2}$$

where α is the pixel density factor, varied in steps of 0.1. This formulation ensures that regions with higher $G_{i,j}$ values, which are critical for the downstream task, are assigned higher quality based on their corresponding $Q_{i,j}$. The ROI map is a scaled variant of the saliency (importance) map, with α serving as the density factor to determine how many pixels represent the objects of interest. Figure 3 (left) illustrates the percentage of relevant pixels assigned for varying α values, with 80% representing high density for objects of interest, and the remaining pixels are assigned 20%. On the right, the 3D plot demonstrates the spatial distribution of pixel importance, where the x and y-axes correspond to the pixel grid, and the z-axis represents the importance value $P(x,y)$ of each pixel.

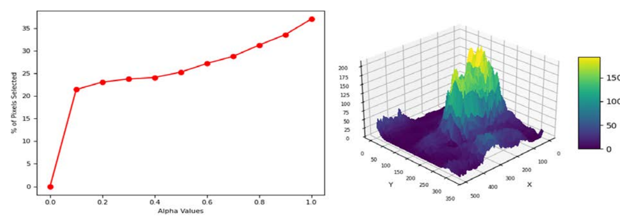


Fig. 3: Percentage of pixels assigned density = 80 for each α: Higher α values allocate more relevant pixels with higher quality (left). Spatial importance distribution for $\alpha = 0.5$ (right)

Channel-aware network. We train a 2-layer Multilayer Perceptron (MLP) with an embedding layer to predict the optimal α value for a given channel condition. This α can be interpreted as a balancing factor between rate and distortion in image compression. Lower α values prioritize rate reduction (more compression), while higher α values focus on minimizing distortion. Our goal is to predict α such that:

$$\text{Minimize: } T_{\text{latency}}, \quad P_{\text{packet loss}}$$
$$\text{Maximize: PSNR (Peak Signal-to-Noise Ratio)}$$

This represents a multi-objective optimization problem, as reducing latency and packet loss typically conflicts with maximizing PSNR. To address this, we train the model on a real-world collected dataset where, given channel conditions such as link bitrate (B) and RSSI (R), the model predicts an optimal α to balance these objectives. Lower α values reduce latency and packet loss by compressing the image more aggressively, while higher α values improve PSNR but increase latency and retransmission risk. To formalize this optimization, we define a composite utility function:

$$U = w_S \cdot S_{\text{n}} + w_L \cdot L_{\text{n}} + w_P \cdot P_{\text{n}} \tag{3}$$

where:

$$S_{\text{n}} = \frac{S - S_{\min}}{S_{\max} - S_{\min}}, L_{\text{n}} = \frac{L_{\max} - L}{L_{\max} - L_{\min}}, P_{\text{n}} = \frac{P_{\max} - P}{P_{\max} - P_{\min}} \tag{4}$$

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on July 03,2025 at 16:27:53 UTC from IEEE Xplore. Restrictions apply.

The weights w_S, w_L, and w_P determine the relative importance of PSNR, latency, and packet loss, respectively, and sum up to 1. The model is trained to maximize this utility function, ensuring that the predicted α value minimizes latency and packet loss while maximizing PSNR under varying channel conditions.

Alpha-conditioned image reconstruction at the receiver. For α-conditioned image reconstruction, we train a DNN to predict the reconstructed image given a compressed input and its associated alpha value, which governs the compression intensity. The model comprises three main components: ① a convolutional encoder to extract feature maps from the compressed image, ② a fully connected (FC) network to process the scalar alpha value, and ③ a Feature-wise Linear Modulation (FiLM) [21] layer to dynamically modulate the image features based on the alpha value. The FiLM operation is formulated as:

$$\text{FiLM}(F,\alpha)=\gamma(\alpha)\cdot F+\beta(\alpha), \tag{5}$$

where $F\in\mathbb{R}^{C\times H\times W}$ represents the image feature maps, and $\gamma(\alpha),\beta(\alpha)\in\mathbb{R}^{C}$ are scaling and shifting parameters derived from the alpha value via the fully connected network. The modulated features are passed through a decoder network to generate the reconstructed image. The training process minimizes the mean squared error (MSE) between the reconstructed image and the ground truth original image. By incorporating the alpha value as a conditioning parameter, the model learns to adapt the reconstruction process based on the compression intensity.

2) Inference Phase

After training the downscaling, task-aware, and channel-aware networks at the transmitter and the alpha-conditioned image reconstruction network at the receiver, the models are quantized in real time based on the available resources to ensure efficient deployment. To determine the appropriate quantization level, we compute a Dynamic Resource Efficiency Index (R) as a weighted combination of the normalized available memory and signal strength for both the transmitter and receiver. This score is defined as:

$$R=w_m\cdot\frac{\text{Available Memory}}{\text{Total Memory}}+w_s\cdot\frac{\text{Current Signal Strength}}{\text{Maximum Signal Strength}}, \tag{6}$$

where w_m and w_s are weights that represent the relative importance of memory and signal strength, $w_m+w_s=1$. We assigned both w_m and w_s a value of 0.5. Based on R, the quantization level is dynamically chosen as:

$$b=\begin{cases}\text{FP32}, & \text{if } R\geq x,\\ \text{FP16}, & \text{if } y\leq R<x,\\ \text{INT8}, & \text{if } R<y,\end{cases}$$

where x and y are predefined thresholds that determine the quantization level. For high resource availability ($R\geq x$), FP32 precision is selected for maximum accuracy. For moderate resources ($y \leq R < x$), FP16 is used to balance accuracy and efficiency. For low resource availability ($R < y$), INT8 quantization is employed to minimize memory and computational overhead. This approach ensures that the model adapts dynamically to varying memory and bandwidth conditions, optimizing performance without exceeding resource constraints. After quantizing the models, *TAVIC-DAS* further performs all the steps as highlighted in the System Overview and illustrated in Fig 2. Algorithm 1 discusses the steps in detail.

Algorithm 1 *TAVIC-DAS* Transmitter and Receiver Operations

Transmitter Operation ⋆
Input: Image $X\in\mathbb{R}^{m\times n}$
Dynamic Model Quantization:
Compute R and assign quantization ▷ Equation 6
Task and Channel-Aware Compression:
Resize image to $X_{\text{downscaled}}\in\mathbb{R}^{m'\times n'}$ ▷ Sec 4.1
Generate saliency map $G\in\mathbb{R}^{m'\times n'}$ ▷ Sec 4.1
Obtain channel conditions (B,R)
Predict optimal α using channel-aware network ▷ Sec 4.1
Compute ROI map $Q_{i,j}=\alpha\cdot G_{i,j}$ ▷ Equation 2
Compress image using ROI map Q
Transmission:
Transmit compressed image and α to receiver
Receiver Operation ⋆
Input: Compressed image, α
Dynamic Model Quantization:
Compute R and assign quantization
Image Reconstruction:
Reconstruct image using α-conditioned reconstruction network
Downstream Processing:
Perform instance segmentation on reconstructed image

V. Experimental Details

A. Data Collection

The streaming modality employed in this study is image data. To train the model for predicting alpha values under varying channel conditions, a comprehensive dataset was collected in real-time encompassing both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios. Data was gathered using two Doodle Labs Mesh Rider Radios operating at 915 MHz with a bandwidth of 20 MHz. Data collection was conducted across three distinct Modulation and Coding Scheme (MCS) rates: *MCS0, MCS2, and MCS4*. We automated the image transmission process by developing a bash script that utilizes Rsync. Successful data traversal through the radio interface was verified using Wireshark packet capture analysis. During each Rsync transmission loop, RSSI values were retrieved from the radios via command-over SSH, and retransmission metrics were recorded using the Linux tool iPerf. The image dataset comprises streamed images from a combination of PASCAL-VOC and Cityscapes. A preview of the collected data samples to train the channel-aware module to output α is presented in Table I, showcasing key metadata attributes. We showcase the data for one single image. However, a set of images was used to train and test the models.

B. Model Hyperparameters

We train the α-conditioned image reconstruction network with a batch size (B) of 16 and a learning rate (LR) of 0.001 for 50 epochs (E). For the downscaling network, training is performed with B=8, and LR=0.001 for E=100 using Cross Entropy Loss and the Adam optimizer. We use the same LR as above for the channel-aware network and train it for E=100.

C. Device Implementation Details

We deployed *TAVIC-DAS* on two platforms: a robotic device (*ROSMaster X3*) and an edge device (*Jetson Xavier*). To

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on July 03,2025 at 16:27:53 UTC from IEEE Xplore. Restrictions apply.

Alpha	FileSize (bytes)	Latency (ms)	Retransmission	RSSI
0.0	29703	422	2	-59
0.1	38167	443	2	-58
0.2	39016	452	3	-58
0.3	39409	567	2	-59
0.4	39653	585	3	-59
0.5	40135	605	3	-59
0.6	40877	623	3	-57
0.7	41493	684	3	-59
0.8	42476	700	4	-56
0.9	43250	760	4	-60
1.0	44265	832	4	-60

TABLE I: Training samples to output α for a particular image and bitrate using Doodle Radio

evaluate *TAVIC-DAS* for long-range communication, we utilized Doodle Labs Mesh Rider Radios (model RM-1700-22M3) operating at a 915 MHz frequency, commonly used by LoRa. The radios were integrated with the devices to facilitate data transmission over the 915 MHz band.

VI. Result Analysis

We evaluate *TAVIC-DAS* against baselines, (a) QMAP-SFT [10] and AID [19], in terms of reconstruction quality, resource efficiency (computation and communication), and performance on the downstream task (instance-segmentation).

A. Baseline Comparison

Reconstruction Quality Evaluation. We evaluated the reconstruction quality of images at the receiver using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) as metrics (Figure 4). *TAVIC-DAS* achieves the highest SSIM of 0.91, outperforming QMap-SFT by 3% and AID by 5.56%, demonstrating its superior structural preservation even when transmitting only the alpha value and not the latent features of the ROI map. This highlights the efficiency of *TAVIC-DAS* in bandwidth-constrained scenarios while maintaining high-quality reconstruction. Furthermore, *TAVIC-DAS* achieves a competitive PSNR, which is only marginally lower than QMap-SFT but significantly higher than AID, demonstrating its ability to balance computational efficiency with reconstruction quality. We also compared *TAVIC-DAS* with the widely used (but task-unaware) codec JPEG XL, which achieved 30.25 dB PSNR and 0.85 SSIM, which is 4.69%↓ in PSNR and 6.59%↓ in SSIM compared to *TAVIC-DAS*.

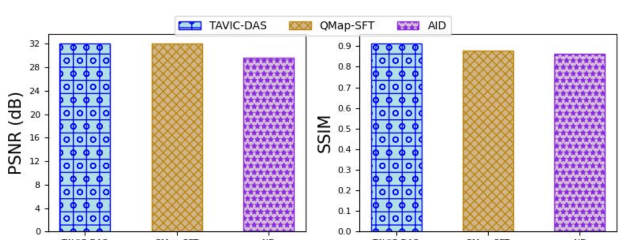


Fig. 4: Comparison of reconstruction quality across baselines

Latency and Power Consumption Analysis

We compare *TAVIC-DAS* with QMap-SFT and AID to evaluate total latency and power consumption, where total latency encompasses compression, transmission, and reconstruction. As shown in Figure 5, *TAVIC-DAS* achieves a lower total latency of 3.22s compared to QMap-SFT's 4.073s. This reduction is primarily due to *TAVIC-DAS* 's significantly faster transmission time (0.3s) versus QMap-SFT's 0.6s, as *TAVIC-DAS* transmits only α and not the quantized latent representation. This demonstrates *TAVIC-DAS* 's efficiency in achieving faster and more effective image reconstruction. We observe that AID achieves the lowest latency of 3.03s, primarily because it lacks a task-aware component, such as our semantic segmentation block, which supports downstream tasks and adds approximately 0.2s to the latency. Also, it does not account for channel conditions, thus increasing the latency in low-bandwidth conditions. In contrast, *TAVIC-DAS* shows significant advantages over AID, achieving a 7.62%↑ improvement in PSNR and a 5.56%↑ increase in SSIM, highlighting its superior reconstruction quality. *TAVIC-DAS* demonstrates improved energy efficiency, consuming 11.10W compared to 12.68W with QMap-SFT. Power consumption of AID is lower due to the absence of a task-aware module at the sender, which is essential for image reconstruction.

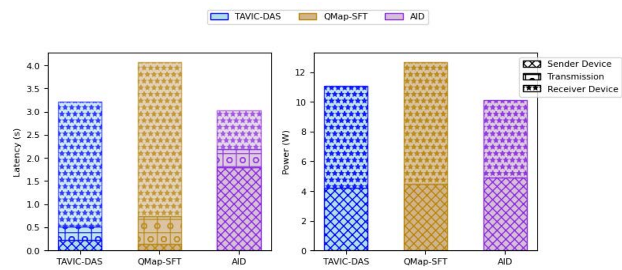


Fig. 5: Comparison based on latency and power consumption

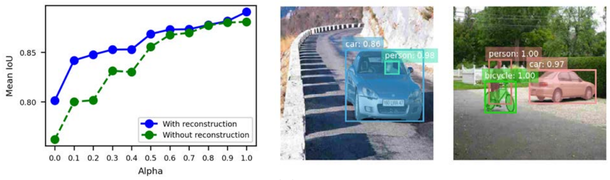
B. Instance Segmentation Analysis

We employ Mask R-CNN [22] for instance segmentation, aiming not only to perform segmentation at the receiver side but also to analyze the impact of α on segmentation performance. To this end, we conduct a detailed study by tracking the Intersection-over-Union (IoU) over our testing dataset for all α values. Fig 6a demonstrates the Mean IoU for each of the images with and without the reconstruction module for varying α. While the performance of the model is comparable with and without reconstruction for higher α values ($\alpha = 0.9$) (IoU 0.887) with reconstruction and (IoU 881) without reconstruction, it is clear that reconstruction enhances the accuracy with lower α values ($\alpha = 0.1$) (IoU 0.841) with reconstruction and (IoU 0.8) without reconstruction. Upon further investigation, we observed that $\alpha = 0.5$ and $\alpha = 0.6$ perform comparably to $\alpha = 1.0$ in terms of IoU for images with fewer instances. However, as the number of instances increases, higher α values begin to show superior performance (refer to figure 6b). Thus, an additional module at the receiver incorporating a choice of reconstruction will enhance the computational efficiency of the devices, which we shall delve into in the future.

C. Ablation Study

We perform a comprehensive ablation study (table II) to examine the impact of *TAVIC-DAS* 's key components on ① PSNR (dB), ② SSIM, ③ Latency (s), and ④ Power consumption (W). The components sequentially removed to analyze their impact are ① Task-aware network, ② Downscaling network, and ③ Alpha-conditioned reconstruction

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on July 03,2025 at 16:27:53 UTC from IEEE Xplore. Restrictions apply.



(a) IoU vs. α with & without reconstruction.

(b) Instance segmentation on reconstructed images: (left) optimal $\alpha = 0.5$, (right) optimal $\alpha = 1.0$ as number of objects increase (right).

Fig. 6: Results for instance segmentation using *TAVIC-DAS*

network. Removing the α-conditioned reconstruction network results in a decline in PSNR to 29.88 (6.87% $\downarrow$) and SSIM to 0.8916 (2.34% $\downarrow$), highlighting its critical role in enhancing the quality of reconstructed images. As anticipated, this also leads to reduced latency (3.14s) and lower power consumption (11.05W). Excluding the downscaling network results in a slight increase in SSIM to 0.9141; however, the latency is increased to 3.19s. Removing the task-aware network leads to a reduction in both PSNR and SSIM, which drop to 29.32 and 0.8894, respectively. The latency and power consumption also decrease from 3.22s to 3.02s and from 11.19W to 11.05W, respectively. The integration of all these components demonstrates *TAVIC-DAS* 's capability to efficiently reconstruct images while maintaining low latency, and reduced power consumption.

Task-aware network	Downscaling network	α-conditioned network	PSNR (dB)	SSIM	Latency (s)	Power (W)
✓	✓	✓	32.01	0.9130	3.22	11.19
✓	✓	✗	29.88	0.8916	3.14	11.05
✓	✗	✓	32.00	0.9141	3.19	11.10
✗	✓	✓	29.32	0.8894	3.02	11.05

TABLE II: Ablation study of *TAVIC-DAS* on PASCAL VOC

VII. Conclusion and Future Work

We proposed *TAVIC-DAS* , a novel image compression framework designed to deliver high-quality reconstructed images while being highly effective in low-bandwidth environments. By incorporating both network-channel parameters and downstream task-specific information, *TAVIC-DAS* achieves higher PSNR $\approx$ 7.62% compared to state-of-the-art techniques. The distributed task execution and model quantization help improve the resource efficiency by $\approx$ 6.39%. This highlights *TAVIC-DAS* as a versatile and robust solution for resource-constrained scenarios, paving the way for future advancements in task-aware and channel-adaptive image compression frameworks. In the future, we aim to extend this framework to other modalities, including video compression and 3D LiDAR point clouds involving multiple agents. We also intend to explore various network parameters in the future.

VIII. Acknowledgement

This work has been partially supported by ONR Grant #N00014-23-1-2119, U.S. Army Grant #W911NF2120076, NSF CAREER Award #1750936, NSF REU Site Grant #2050999, NSF CNS EAGER Grant #2233879, and U.S. Army Grant #W911NF2410367.

References

[1] Arian Bakhtiarnia, Błażej Leporowski, Lukas Esterle, and Alexandros Iosifidis. Analysis of the effect of low-overhead lossy image compression on the performance of visual crowd counting for smart city applications. In *2022 IEEE International Smart Cities Conference (ISC2)*, pages 1–5, 2022.

[2] Shohei Uchigasaki, Tomo Miyazaki, and Shinichiro Omachi. Deep image compression using scene text quality assessment. *Pattern Recogn.*, 142(C), October 2023.

[3] Zixiang Xiong, Angelos D. Liveris, and Samuel Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21:80–94, 2004.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Commun. ACM*, 63(11):139–144, October 2020.

[5] Lingling Pu, Michael W. Marcellin, Ali Bilgin, and Ami Ashok. Image compression based on task-specific information. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4817–4821, 2014.

[6] Ching-Chung Wei, Shu-Ting Chen, and Pei-Yi Su. Image transmission using lora technology with various spreading factors. In *2019 2nd World Symposium on Communication Engineering (WSCE)*, pages 48–52, 2019.

[7] Xu Kang, Bin Song, Jie Guo, Zhijin Qin, and Fei Richard Yu. Task-oriented image transmission for scene classification in unmanned aerial systems. *IEEE Transactions on Communications*, 70(8):5181–5192, 2022.

[8] Yi Ma, Yongqi Zhai, Chunhui Yang, Jiayu Yang, Ruofan Wang, Jing Zhou, Kai Li, Ying Chen, and Ronggang Wang. Variable rate roi image compression optimized for visual quality. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1936–1940, 2021.

[9] Chia-Hao Kao, Ying-Chieh Weng, Yi-Hsin Chen, Wei-Chen Chiu, and Wen-Hsiao Peng. Transformer-based variable-rate image compression with region-of-interest control. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 2960–2964, 2023.

[10] Myung-Sin Song, Jinyoung Choi, and Bohyung Han. Variable-rate deep image compression through spatially-adaptive feature transform. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2360–2369, 2021.

[11] Jinyoung Choi and Bohyung Han. Task-aware quantization network for jpeg image compression. In *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX*, page 309–324, Berlin, Heidelberg, 2020. Springer-Verlag.

[12] Heewon Kim, Myungsub Choi, Bee Lim, and Kyoung Mu Lee. Task-aware image downscaling. In *European Conference on Computer Vision*, 2018.

[13] Maxime Kawawa-Beaudan, Ryan Roggenkemper, and Avideh Zakhor. Recognition-aware learned image compression. In *Computational Imaging*, 2022.

[14] Uri Gadot, Assaf Shocher, Shie Mannor, Gal Chechik, and Assaf Hallak. Real time macro-block rate control for task-aware video compression using reinforcement learning, 2024.

[15] Xingtong Ge, Jixiang Luo, Xinjie Zhang, Tongda Xu, Guo Lu, Dailan He, Jing Geng, Yan Wang, Jun Zhang, and Hongwei Qin. Task-Aware Encoder Control for Deep Video Compression . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26036–26045, Los Alamitos, CA, USA, June 2024. IEEE Computer Society.

[16] Khizar Anjum, Zhile Li, and Dario Pompili. Acoustic channel-aware autoencoder-based compression for underwater image transmission. In *2022 Sixth Underwater Communications and Networking Conference (UComms)*, pages 1–5, 2022.

[17] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *ArXiv*, abs/1706.05587, 2017.

[18] Mark Sandler, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[19] Hangyul Choi, Seongmoon Jeong, Sangwoon Kwak, Soon-Heung Jung, and Jong Hwan Ko. Adaptive image downscaling for rate-accuracy-latency optimization of task-target image compression. In *2024 IEEE 6th International Conference on AI Circuits and Systems (AICAS)*, pages 347–351, 2024.

[20] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[21] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: visual reasoning with a general conditioning layer. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018.

[22] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn, 2018.

Authorized licensed use limited to: University of Maryland Baltimore Cty. Downloaded on July 03,2025 at 16:27:53 UTC from IEEE Xplore. Restrictions apply.