Diff-GO: Diffusion Goal-Oriented Communications to Achieve Ultra-High Spectrum Efficiency

Achintha Wijesinghe[1], Songyang Zhang[2], *Member, IEEE*, Suchinthaka Wanninayaka[1], Weiwei Wang[1], and Zhi Ding[1], *Fellow, IEEE*
[1]University of California, Davis, CA, USA
[2]University of Louisiana at Lafayette, LA, USA

***Abstract*—The latest advances in artificial intelligence (AI) present many unprecedented opportunities to achieve much improved bandwidth saving in communications. Unlike conventional communication systems focusing on packet transport, rich datasets and AI makes it possible to efficiently transfer only the information most critical to the goals of message recipients. One of the most exciting advances in generative AI known as diffusion model presents a unique opportunity for designing ultra-fast communication systems well beyond language-based messages. This work presents an ultra-efficient communication design by utilizing generative AI-based on diffusion models as a specific example of the general goal-oriented communication framework. To better control the regenerated message at the receiver output, our diffusion system design includes a local regeneration module with finite dimensional noise latent. The critical significance of noise latent control and sharing residing on our Diff-GO is the ability to introduce the concept of "local generative feedback" (Local-GF), which enables the transmitter to monitor the quality and gauge the quality or accuracy of the message recovery at the semantic system receiver. To this end, we propose a new low-dimensional noise space for the training of diffusion models, which significantly reduces the communication overhead and achieves satisfactory message recovery performance. Our experimental results demonstrate that the proposed noise space and the diffusion-based generative model achieve ultra-high spectrum efficiency and accurate recovery of transmitted image signals. By trading off computation for bandwidth efficiency (C4BE), this new framework provides an important avenue to achieve exceptional computation-bandwidth tradeoff.**

***Index Terms*—Goal-oriented, diffusion model, generative AI, local generative feedback, computation for bandwidth efficiency.**

I. Introduction

In the ever-evolving landscape of communications and networking, novel technologies, such as the sixth generation (6G) [1] and autonomous driving, are expected to emphasize the need for effective and efficient methods of spectrum utilization by incorporating the rapidly growing artificial intelligence (AI). In recent years, the field of generative AI has witnessed a paradigm shift with the advent of diffusion models [2]. These models, built upon the foundations of deep learning and neural networks, have revolutionized the way we process and understand language and visuals ushering in a new era for bandwidth-efficient communication. These advances present opportunities for the study of semantic communication [3] by focusing on the delivery of a message's "meaning" rather than the original signal manifestation. Semantics-based communication has proven to be a promising direction toward goal-oriented communication by having a plethora of far-reaching real-world applications with the prevailing computational-sound smart transmitters. For example, a system to communicate text information semantically with a customized loss function was proposed in [4]. Other applications also include UAV communication, remote image sensing and fusion, intelligent transportation, and healthcare [5].

The *goal* of transferring semantic meaning in communication is only a special case of a general framework of *goal-oriented* communications. Semantic communications assume that the goal is to deliver the meaning rather than the exact fully reconstructed message [6] to the recipients, thereby enabling efficient information transfer. However, in a more broad context, the designers are more interested in what the end goal of information transfer is, beyond the mere semantic meaning. For example, if the goal of data collection is to facilitate the model training of autonomous driving, the autonomous vehicle (AV) understands its surroundings by processing images from networked cameras, where information such as the depth, direction, or size of surrounding objects are more critical to AV decision making than the colors or the brands of adjacent cars. Therefore, in this goal-oriented communication example, full resolution image is redundant, whereas a simple semantic description of the camera scene is insufficient. Generally, goal-oriented communications rely on local computation engine to determine what critical information should be transmitted and at what accuracy. Bandwidth efficiency can be gained from computation on the transmitter side to outperform classical communication approaches. In many applications including autonomous driving, power transceivers with sufficient computation power is affordable to achieve the desired computation-bandwidth tradeoff without compromising the specific downstream goals for communications, such as decision making. The improved efficiency, link reliability, user quality of experience, and smoother cross-protocol communications justify the strong role of computation for bandwidth efficiency (C4BE) designs in future communications.

Some recent works have started to consider AI-enabled frameworks, integrating many concepts in AI, such as machine learning, causal reasoning, and minimum description length theory [7]. One interesting diffusion-based encoder proposed in [8] aims to combat channel noise by sharing a pre-trained diffusion model is used for information generation. Modern wireless systems, however, rely on techniques such as hybrid

ARQ to combat packet errors or losses and do not involve noise effects directly onto the raw messages. Another work uses masked vector quantized-variational autoencoder (VQ-VAE) for a goal-oriented communication setup [9]. Other typical AI-enabled communication schemes also include auto-encoders [10] joint task and data-oriented semantic communications [11], semantic multi-modal data systems [12], and semantic channel coding [13]. Due to the page limits, interested readers could refer to [1], [6] for a more complete survey on semantic communications as a special case of Goal-Oriented (GO) communication systems.

Despite the successes, the existing approach to semantics-based frameworks to enhance bandwidth utilization faces some obstacles. Firstly, the structure of conventional AE-based architecture limits the flexibility and ability of semantic re-generation, where its integration with the emerging generative learning models might be more informative and effective. Secondly, the data link is often designed in one direction and no consideration is given for quality of service (QoS) control, especially for diffusion-based architecture [8]. Due to the random nature of diffusion models, sharing constraint conditions alone has no control over the generated output accuracy. As a result, the transmitter is unaware of the reconstructed output on the receiver side. On the other hand, using a random noise latent to generate the output may not guarantee the closest representation of the input information, especially in images. Moreover, the questions of how to reduce the communication overhead and how to fully leverage the flexibility of generative learning models remain open.

To tackle the aforementioned practical issues in existing designs, we propose a novel *Diff*usion-based, *G*oal-*O*riented (Diff-GO) communication framework utilizing generative AI with local generative feedback (local-GF), where a new lightweight low dimensional noise space is proposed for the training of diffusion models. Our proposed Diff-GO communication framework aims to reduce the communication overhead and implement effective information regeneration to satisfy the required goal-oriented QoS (GO-QOS)at the receiver by designing a novel local generative feedback (local GF) at the transmitter. Our contribution can be summarized as follows:

- We propose a novel C4BE design principle and a Diff-GO communication architecture. Our system design employs local GF for GO-QOS control.
- We propose a new training approach for diffusion models. Specifically, we introduce a noise space mapping which enables noise latent sharing between the transmitter and the receiver at a very low communication cost.
- Through rigorous tests, we demonstrate the efficacy of Diff-GO as an effective C4BE design.

II. Overall Architecture

In this section, we first present our novel communication architecture for task-specific semantics based on the diffusion model as illustrated in Fig. 1.

A. Exemplary Application in Autonomous Driving

In this work, as an example, we consider autonomous driving as an exemplary application and focus on the transmission of city street images for "smart driving". To ensure the collaboration and federation among different autonomous cars in smart driving, efficient communication on street traffic conditions collected from cameras or sensors plays an important role in identifying objects and making safe decisions. In particular, it is paramount to capture the correct road signs and patterns in autonomous driving, where the edge map and segmentation map contain rich go-oriented (GO) information for driving intelligence. For convenience, we will introduce our overall Diff-GO communication architecture in the background of autonomous driving. Note that, beyond autonomous driving, our proposed framework can be easily deployed in any Diff-GO communication scenario with suitable use-case conditions.

B. Deployment and Local Generation At Transmitter Side

To leverage generative learning in Diff-GO communication, we investigate the utilization of diffusion models. For communication efficiency, we propose a novel quantized noise space rather than the continuous noise space spanned from the classical forward diffusion, inspired by the idea of introducing a local generative to eliminate any ambiguity introduced by the nature of the Diff-GO. In many proposed semantic communication frameworks such as [8], the information generated on the receiver side is unknown to the transmitter. To alleviate this issue, we propose a novel low-dimensional (low-DIM) noise space spanned by a linear combination of known n number of seeds and share the corresponding weights (w) with the receiver.

As depicted in Fig. 2, the transmitter is equipped with a pre-trained diffusion model. The diffusion model is pre-trained on a pre-selected noise basis spanned from a known set of n number of seeds. We will elaborate on the low-DIM noise space in Section III-C. Our transmitter pipeline starts with noise generation. We first utilize classical forward diffusion to derive the noise latent of any given image. With the derived noise latent, we project the noise latent to the low-DIM noise space and find the optimal weight vector $\mathbf{W}$, which will be presented in Section III-C in detail. Simultaneously, we extract conditions from the given image. For a task such as autonomous driving, we propose to use a segment map and edge map to enhance the generation quality of the diffusion model.

To further reduce the communication overhead, we propose a hierarchical approach to send the most significant n_i weights $n_1 < n_2 <, \ldots, n_p < n$ to the receiver. These $n_1, \ldots, n$ are predefined. Subsequently, we iterate through each n_i and generate a low-DIM noise representation of the original noise latent by linearly combining the selected n_i weights with the corresponding noise basis generated from seeds. Next, the low-DIM noise latent is fed through the diffusion model for denoising by conditioning on the extracted condition. The output from the diffusion model is sent for evaluation. The evaluation is done by our proposed local generative feedback

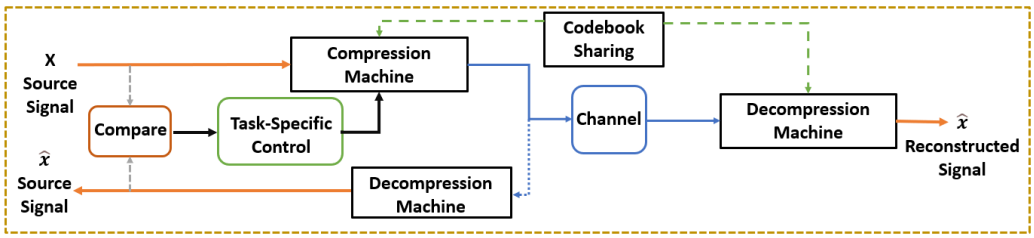

Fig. 1: Proposed semantic communication framework: The overall diagram;

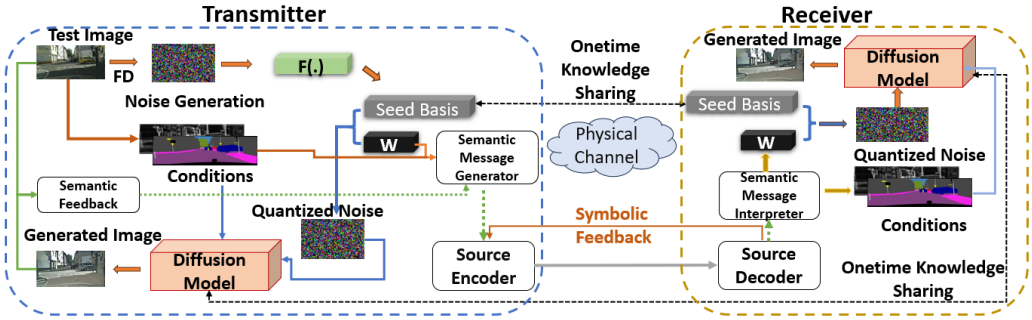

Fig. 2: Proposed semantic communication framework: The detailed mechanism.

block. We further illustrate this block in Section III-D. If the GO-QOS score is adequate for the downstream task, we pick the corresponding weights for the diff-GO message generation followed by transmitter-level encoding and communicate with the receiver through the physical channel.

C. Data Transmission and Regeneration at Receiver Side

As in any classical communication system, we can apply feedback and error correction schemes to evaluate the validity of the received information. We adopt a semantic message interpreter on the receiver side to uncover the sent $\mathbf{W}$ and diffusion conditions. We combine the seed basis and $\mathbf{W}$ linearly to recover the best noise latent and feed it to the diffusion model along with the diffusion conditions. This process guarantees to generate the exact image generated on the transmitter side.

III. Methodology

A. Diffusion Models

Diff-GO pivots around goal-oriented diffusion models. In this part, we first briefly introduce the structure of diffusion models [2]. In general, the training of a diffusion model consists of two processes: 1) forward diffusion and 2) backward diffusion. Let any data point be denoted by $x_0 \sim q(x_0)$ and $\{x_1, \cdots, x_T\}$ represent latent with similar dimension to x_0.

In the forward diffusion, a given data point is used to learn a noisy latent by iterative noise addition. In this process, the posterior $q(x_{1:T}|x_0)$ is determined as a Markov chain that gradually adds Gaussian noise to the given data point. For example, in autonomous driving, noises will be added to the street image from cameras at each step. A variance schedule $\{\beta_1, ..., \beta_T\}$ is used at each iteration [2] as follows.

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}), \tag{1}$$

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathcal{I}). \tag{2}$$

In the backward diffusion process, a neural network is utilized as a denoising auto-encoder to learn the added noise in each forward process to characterize the joint distribution $p_\theta(x_{0:T})$. The reverse process is modeled as a Markov chain and the aforementioned auto-encoder is expected to learn the Gaussian transitions. This transitions starts at $p(x_T) = \mathcal{N}(x_T; \mathbf{0}, \mathcal{I})$ and follows the tranisition steps below.

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t), \tag{3}$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \nu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \tag{4}$$

B. Go-Oriented Diffusion Model

The promise of low communication cost go-oriented (GO) communication is viable due to the amazing properties of diffusion models. In this part, we introduce the design of Diff-GO communication systems.

Since diffusion models provide the freedom of conditioning using GO-QOS, we could modify the reverse process (backward diffusion) with a given condition y in conditional diffusion models, recalculated by,

$$p_\theta(x_{0:T}|y) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t, y), \tag{5}$$

$$p_\theta(x_{t-1}|x_t, y) = \mathcal{N}(x_{t-1}; \nu_\theta(x_t, y, t), \Sigma_\theta(x_t, y, t)). \tag{6}$$

As introduced in the work [8], it is possible to share communication-friendly conditions with the receiver side and generate a possible GO representation of the input data, such as the city images in autonomous driving, by starting with any random noise. However, such an approach has no guarantee that the reconstructed image is accurate in terms of GO-QOS since some sampled noise latent vectors of the learned data latent may not map back to the original space and fail to represent the original image distribution, resulting in generating an image without any valuable information. Fortunately, a conditional diffusion model provides 2 degrees-of-freedom to

Fig. 3: Our receiver model output.

Fig. 4: Output from the system of [8]

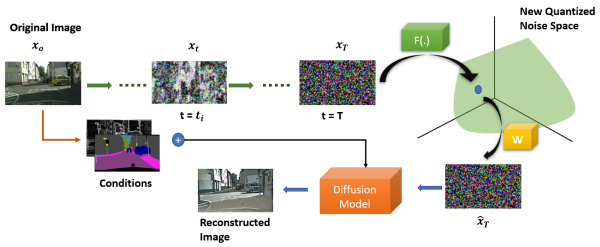

Fig. 5: Proposed Forward Diffusion: We perform classical forward diffusion up to $t = T$ and map x_T to a new space spanned by N number of known noise vectors. These vectors are defined by a seed value (a numerical value). The diffusion model is then trained on the new projected noise space

alter the regeneration to ensure GO-QOS consistency: 1) alter the generated image by changing the input noise latent; or 2) use the conditions to manipulate the model output.

To leverage the critical information, we use the edge map as conditions, generated using Canny edge detection from the segmentation map. Compared to the generated images in the [8] which only utilize the segmentation map of street images, we can preserve essential information such as road signs by introducing edge maps as conditions, as depicted in Fig 3 and Fig. 4. Despite the successes of the semantic diffusion model, only sharing the conditions with the receiver still cannot guarantee the quality of regenerated data. Moreover, the transmitter side cannot be aware of the reconstructed output on the receiver side without any feedback. To alleviate these downsides, we now introduce local generative feedback with a proper choice of noise latent vectors next.

Algorithm 1: Diff-GO Communication Training

Input: Training data: $\mathcal{D}$
Input: Number of vectors in noise basis: N
Output: Trained diffusion model: D
Output: Seed set: $\mathcal{S}$

Initialize $\mathcal{S}$ randomly: $\mathcal{S} \leftarrow \{s_1, \ldots, s_N\}$
while *D is not converged* **do**
  for *each* $d \in \mathcal{D}$ **do**
    $x_t \leftarrow$ output of FWD(d)
    $m_t \leftarrow$ GO-QOS for d
    if $t == T$ **then**
      Solve Eq. (7) $w \leftarrow F(x_T, \mathcal{S})$
      $\hat{x_T} \leftarrow \sum_{i=1}^{N} w_i \times s_i$
      $x_t \leftarrow \hat{x_T}$
    Train D: $D \leftarrow$ BWD(x_t, m_t)
Share D and $\mathcal{S}$ with the receiver

C. Noise Quantization

To ensure accurate feedback, we first need to identify the best possible noise latent for a given image. Here, We argue that the best possible noise latent for any given image shall be characterized by the forward diffusion process. This is because the diffusion process is an encoder-decoder structure, where the forward diffusion encodes a given image to a noise latent of the same dimensions of the image, and the backward diffusion decodes or denoises it. Therefore, intuitively, a well-trained diffusion model is capable of denoising a noise latent derived from the forward diffusion back to the original image distribution. Therefore, at the inference, we propose to use the forward diffusion to map a given image to its noise latent. However, if we have to share the derived noise directly with the receiver side, it serves no good since the noise is not compressible. To address this, we propose a new mapping from the noise space of the forward diffusion to a predefined new noise space. Let $s_1, s_2, \cdots, s_n$ be randomly selected n number of noise seeds and $N_1, N_2, \cdots, N_n$ be the corresponding noise latents generated by each noise seed from a Gaussian distribution. We define a new quantized linear noise space $\mathcal{N}_q$ by the span denoted as $Span(N_1, N_2, \cdots, N_n)$.

As illustrated in Fig 5, we first follow a typical forward and backward diffusion process except for the case when the step $t = T$. Every time, the step hits $t = T$ we learn a function $F(.)$ which maps the noise latent at $t = T$ (N_T) to noise space $\mathcal{N}_q$. i.e., $F : N_T \rightarrow \hat{N_T} \in \mathcal{N}_q$.

Suppose that $w_1, \cdots, w_n$ represents a set of weights. The function $F(.)$ can be obtained as follows.

$$F = \underset{w_1, \cdots, w_n}{\arg\min} ||N_T - \sum_{i=1}^{n} w_i \times N_i||^2 \tag{7}$$

We solve the above optimization problem for each N_T starting with the same weights $w_1, \cdots, w_n$ and use gradient descent to solve the optimization problem in Eq. (7). This approach allows us to train our diffusion model in a known quantized noise space. The main advantage of this approach is that we can use the forward process to derive the best possible noise and represent it by only n number of weights which is a highly compressed representation of the noise latent. Then, the weights can be sent to the receiver side for reconstruction of the exact noise for backward diffusion.

D. Local Generative Feedback

With the quantized space, we can represent the noise with respect to a known basis, where we have the freedom of representing any derived noise latent in a highly compressed representation using n floating points. The new representation therefore allows us to generate the exact information that the receiver is going to generate even before sharing the information with the receiver. This allows us to validate the reconstruction with the original image using goal oriented QoS (GO-QOS) metrics, such as Fréchet inception distance (FID) score [14], Learned Perceptual Image Patch Similarity (LPIPS) [15], Segment Anything Model (SAM) score [16], or any downstream task, which will be further discussed in

Section IV. This approach significantly differs from existing semantic feedback since we do not wait for the receiver reports (e.g. car crash). Therefore, it provides communication efficiency. Agorithm 1 and Algorithm 2 summarize the training and the inference of the proposed method.

IV. Numerical Experiments

We now present the performance of our Diff-GO communication system against existing works. We evaluate all the models on the Cityscape dataset [17] for autonomous driving.

Algorithm 2: Message Inference

Input: Inference image: $\mathcal{I}$
Input: GO-QOS Threshold: τ
Input: GO-QOS metric: $\mathcal{M}$
Output: Weights set w

Initiate a set $\mathcal{P}$ with the different number of important basis vectors to consider $\mathcal{P} \leftarrow n_1, n_2, \ldots, n_q, n$
$x_t \leftarrow$ output of FWD($\mathcal{I}$)
$m_t \leftarrow$ meaning of $\mathcal{I}$
Solve Eq. (7) $w \leftarrow F(x_T, \mathcal{S})$
for *each* $n_i \in \mathcal{P}$ **do**
    $wT \leftarrow$ sort from max to min ($|w|$)[n_i]
    $\hat{w} \leftarrow$ [0 for w_i in w if $|w_i| < wT$ else w_i]
    $\hat{x_T} \leftarrow \sum_{i=1}^{n_i} \hat{w}_i \times s_i$
    $x_t \leftarrow \hat{x_T}$
    $I_i \leftarrow$ BWD(x_t, m_t)
    GO-QOS score $\leftarrow \mathcal{M}(\mathcal{I}, I_i)$
    if $GO - QOS\ score \leq \tau$ **then**
        Share $\hat{w}$ and m_t
Share $\hat{w}$ and m_t

A. Quality of Reconstructed Receiver Message

First, we evaluate the quality of the reconstructed images by different methods, such as generative Semantic Communication (GESCO) [8] and original diffusion (OD). In OD, we use guided diffusion [18] and use the semantic map and edge map as the conditions, where we share the noise latent extracted from the forward diffusion process. In diffusion with random noise (RN), we use a random noise latent to generate results without sharing the noise latent derived from the forward diffusion in Diff-GO. The evaluation is done after training all the models in 250000 steps. For the evaluations, we rely on GO-QOS measurement and avoid any pixel-wise measures, such as mean square error (MSE), in view of communication goals. For example, the same car having different colors reports a higher MSE value even though they are the same for decision making. FID score enables us to evaluate the generated images with respect to human inception at the feature level. Lower FID values represent better reconstruction quality. The results are presented Table I From the results, we can see that the model OD has the lowest FID. The reason behind this observation is that we are sharing the entire noise latent with the receiver side. However, as we mentioned, sharing the noise latent is not feasible and the communication cost is even higher compared with image sharing due to the incompressibility of the noise latent. We also observe the effect of the edge map as a condition when we compare it with the higher FID in the GESCO and OD methods. Similar observations can be made for LPIPS as well. For the SAMSR and SAMSS, presented in Table III, DIff-GO 100 and OD perform closely. We see an increment in performances as n increases in Diff-GO. In the last column of Table III, we present the raw (without any compression) number of floating points we need to communicate through the channel for all three methods. Here, C represents the number of floating points in the shared conditions which is common to all the methods. E represents the size of the edge map (binary map), which counts the additional conditions introduced in this work. Compared to OD, Diff-GO can save up to 0.5 M floating points due to novel quantized noise space. The results imply the capacity of our method to reconstruct the same meaningful data while saving GO communication costs. Table I demonstrates the benefits of encoding random vectors via prior-selected bases.

TABLE I: Semmantic similarity of the generated images with different methods evaluated against different metrics: LPIPS and FID. Some of the results are from the work GESCO [8] (Smaller value is better). We present different choices of n for our proposed Diff-GO.

Method	LPIPS↓	FID↓
SPADE [19]	0.546	103.24
CC-FPSE [20]	0.546	245.9
SMIS [21]	0.546	87.58
OASIS [22]	0.561	104.03
SDM [23]	0.549	98.99
OD	0.2191	55.85
GESCO	0.591	83.74
RN	0.3448	96.409
Diff-GO (n=20)	0.3206	74.09
Diff-GO (n=50)	0.2697	72.95
Diff-GO (n=100)	0.2450	68.59

B. Evaluation in Downstream Tasks

Next, we measure the performance of the proposed framework in downstream tasks. We first present the results of the object detection. For the evaluations, we use the pre-trained object detection model from [24]. In this experiment, we evaluate the mean intersection over union (mIoU) for three different objects of interest. From Table II, Diff-GO achieves superior performance similar to OD, while the GESCO underperforms. Note that, here, we use forward diffusion to derive the best noise for all the methods.

We also evaluate the performance in depth map estimation as presented in in Table II. Here evaluate the depth estimation using root mean square error (RMSE) between ground truth and reconstructed image. For depth map generation from street images, we use a pre-trained model from the work [25]. From Table II, Diff-GO performs the best, in comparison with OD and GESCO, which show the promising power of Diff-GO for downstream tasks of depth sensing in autonomous driving.

C. Ablation Study of Different Size of Noise Space

In this part, we experiment with different numbers of noise basis vectors to assess their impact. We use 100 noise vectors and examine FID, LPIPS, SAMSR, and SAMSS scores

for sharing various quantities, denoted as n_i. We choose $n_i = 1, 10, 20, 50,$ and 100. From Table IV, FID and LPIPS scores decrease as we share more basis vectors, while SAMSR and SAMSS scores increase with n. We also demonstrate the flexibility of using downstream tasks (death estimation) for determining the number of weights to share with the receiver which follows the same trend. As shown in Table IV, even sharing just one essential weight can lead to efficient message regeneration and communication cost savings.

TABLE II: Object detection: mean intersection over the union (mIoU↑) of objects of interest. We use pre-trained model form [24] for object detection on the generated images and error of depth estimation in RMSE.

Method	Car	People	Bicycle	Depth (RMSE↓)
GESCO	68.27	62.89	62.26	0.1489
OD	73.52	**68.96**	67.30	0.1183
Diff-GO 100	**73.55**	67.57	**67.74**	**0.1077**

TABLE III: GO-QOS of reconstructed images in terms of SAM score, and transmitted floating points to the receiver end by three different approaches. (SAMSR: SAM score with respect to the ground truth; SAMSS: SAM score with respect to the segmentation map of the ground truth.)

Method	SAMSR↑	SAMSS↑	Transmitted Floating Points↓
GESCO	0.9738	0.9744	C
OD	**0.9941**	0.9839	C+E+0.5M
Diff-GO 100	0.9927	**0.9864**	C+E+100

TABLE IV: How different QoS scores vary with the number of basis vectors of the noise latent. Depth represents the depth estimation as GO-QOS score. Here, $n_1 = 1, n_3 = 10, n_4 = 20, n_5 = 50, n = 100$

Metric	1	10	20	50	100
FID↓	70.252	69.43	69.45	69.46	68.59
LPIPS↓	0.2727	0.2794	0.2793	0.2793	0.2450
SAMSR↑	0.98611	0.9861	0.9861	0.9861	0.9861
SAMSS↑	0.9854	0.9854	0.9854	0.9854	0.9927
Depth (RMSE)↓	0.1078	0.1077	0.1077	0.1077	0.1077

V. Conclusions and Future Works

This work introduces a novel goal-oriented (GO) communication system that generalizes the concept of semantic communications and prioritizes the downstream tasks that rely on the communicated signals to trade computation for bandwidth efficiency (C4BE). The Diff-GO communication design achieves ultra-high bandwidth efficiency by utilizing generative AI at its core. Unlike many semantic communication works using generative AI, our approach considers the goal of communications beyond the basic language model (or semantics). We leverage a forward diffusion process and use a unique low-dimensional noise space for bandwidth reduction. To ensure goal-oriented QoS (GO-Qos), our approach implements local generative feedback without altering any existing communication links and protocols. Future research will explore different noise spaces for improved performance and consider developing metrics for measuring GO-QoS.

References

[1] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 210–219, 2022.

[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 6840–6851.

[3] E. C. Strinati and S. Barbarossa, "6g networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.

[4] Y. Wang, M. Chen, T. Luo, W. Saad, D. Niyato, H. V. Poor, and S. Cui, "Performance optimization for semantic communications: An attention-based reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2598–2613, 2022.

[5] Y. Liu, X. Wang, Z. Ning, M. Zhou, L. Guo, and B. Jedari, "A survey on semantic communications: technologies, solutions, applications and challenges," *Digital Communications and Networks*, 2023.

[6] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.

[7] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next generation semantic communication networks," *arXiv preprint arXiv:2211.14343*, 2022.

[8] E. Grassucci, S. Barbarossa, and D. Comminiello, "Generative semantic communication: Diffusion models beyond bit recovery," *arXiv preprint arXiv:2306.04321*, 2023.

[9] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications with masked vq-vae enabled codebook," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[10] X. Luo, B. Yin, Z. Chen, B. Xia, and J. Wang, "Autoencoder-based semantic communication systems with relay channels," in *2022 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2022, pp. 711–716.

[11] J. Huang, D. Li, C. Huang, X. Qin, and W. Zhang, "Joint task and data oriented semantic communications: A deep separate source-channel coding scheme," *arXiv preprint arXiv:2302.13580*, 2023.

[12] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, and X. Tao, "A unified multi-task semantic communication system for multimodal data," *arXiv preprint arXiv:2209.07689*, 2022.

[13] J. Dai, P. Zhang, K. Niu, S. Wang, Z. Si, and X. Qin, "Communication beyond transmitting bits: Semantics-guided source and channel coding," *IEEE Wireless Communications*, vol. 30, no. 4, pp. 170–177, 2023.

[14] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[15] H. Talebi and P. Milanfar, "Learned perceptual image enhancement," in *2018 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2018, pp. 1–13.

[16] Y. Li, M. Chen, W. Yang, K. Wang, J. Ma, A. C. Bovik, and Y. Zhang, "Samscore: A semantic structural similarity metric for image translation evaluation," *arXiv preprint arXiv:2305.15367*, 2023.

[17] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3213–3223, 2016.

[18] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8780–8794, 2021.

[19] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2332–2341.

[20] X. Liu, G. Yin, J. Shao, X. Wang *et al.*, "Learning to predict layout-to-image conditional convolutions for semantic image synthesis," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[21] Z. Zhu, Z. Xu, A. You, and X. Bai, "Semantically multi-modal image synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5467–5476.

[22] V. Sushko, E. Schönfeld, D. Zhang, J. Gall, B. Schiele, and A. Khoreva, "You only need adversarial supervision for semantic image synthesis," *arXiv preprint arXiv:2012.04781*, 2020.

[23] W. Wang, J. Bao, W. Zhou, D. Chen, D. Chen, L. Yuan, and H. Li, "Semantic image synthesis via diffusion models," *arXiv preprint arXiv:2207.00050*, 2022.

[24] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision*, 2020, pp. 213–229.

[25] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 179–12 188.