Semantic-Aided Image Transmission System with Unequal Error Protection for Next-Generation Communication Networks

Nargis Fayaz †
Dept. of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi 110016, India
Nargis.Fayaz@ee.iitd.ac.in

Aman Shreshtha †
Dept. of Computer Science and Engg.
Indian Institute of Technology Delhi
New Delhi 110016, India
amanshreshtha@cse.iitd.ac.in

Smruti Sarangi
Dept. of Computer Science and Engg.
Indian Institute of Technology Delhi
New Delhi 110016, India
srsarangi@cse.iitd.ac.in

Ranjan K. Mallik
Dept. of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi 110016, India
rkmallik@ee.iitd.ac.in

Brejesh Lall
Dept. of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi 110016, India
brejesh@ee.iitd.ac.in

***Abstract*—Semantic communication (SC) aims to convey the meaning of data instead of focusing on its bit-by-bit reconstruction. SC finds applications in beyond 5G and 6G networks for artificial intelligence-empowered multimedia content delivery. In this paper, we propose a novel semantic-aided autoencoder-based image transmission system that leverages semantic information in the form of the segmentation map of an image. We demonstrate up to 23% and 18% improvement (in terms of mean square error and peak signal-to-noise ratio, respectively) in the quality of the received image with only 2% extra bandwidth over a traditional autoencoder-based image transmission system. The study also explores channel coding strategies for our proposed system. We focus on the intrinsically robust nature of semantic data, as compared to traditional data, to design low-density parity check code, Hamming code, and polar code-based unequal error protection (UEP) schemes. Comparative evaluations between UEP and equal error protection schemes show that while both approaches yield similar performance, UEP schemes are more efficient.**

***Index Terms*—Autoencoder, image segmentation, low-density parity check (LDPC) codes, multimedia transmission system, polar codes, semantic communication (SC), unequal error protection (UEP).**

I. Introduction

As we progress into beyond 5G and 6G networks, we find that there is a need to view traditional communication systems from a new perspective, focusing on semantics. The objective of semantic communication (SC) is to convey the semantic meaning of the message by sharing a common, prior knowledge between the sender and the receiver [1]. Enhancing traditional systems using SC is a necessary step towards achieving fully intelligent networks. An intermediate step towards moving entirely to SC is semantic-aided communication. Semantic-aided communication represents a paradigm that exploits the noise resilience of semantic content to achieve reduced-loss communication links. A comprehensive examination of the principles of SC for diverse information sources such as images, text, and video is given in [2]. The authors in [3] develop a semantic image transmission system. The scope of their work is restrictive, as they use a generative adversarial network (GAN) to regenerate the image at the receiver solely from its segmentation map, omitting other image details. The concept of deep joint source-channel coding (DeepJSCC) for image transmission is presented in [4]. Authors in [5] present a method to enhance wireless image transmission using domain-specific semantic information. Both studies propose end-to-end wireless systems using deep neural networks to directly map image bits to channel input symbols, ignoring traditional physical layer channel coding and modulation techniques.

Previous works use purely SC systems, disregarding syntactic (traditional) data. In real-world applications, blending SC with traditional communication yields a heterogeneous setting where both semantic and syntactic data coexist, offering combined advantages. These two types of data inherently embody varying degrees of significance and error tolerance. Hence, it would be more efficient to implement unequal error protection (UEP) [6] schemes for the two data types rather than assuming uniform information importance as in a conventional communication framework. In our study, we employ the concept of UEP at the physical layer in the context of semantic-aided image transmission. In our proposed approach, the segmentation map of the image is used to improve image quality at the receiver, reconstructed from the latent map of the image using an autoencoder. The primary contributions of this paper can be summarized as follows:

- We propose a semantic-aided autoencoder-based image

†Authors contributed equally

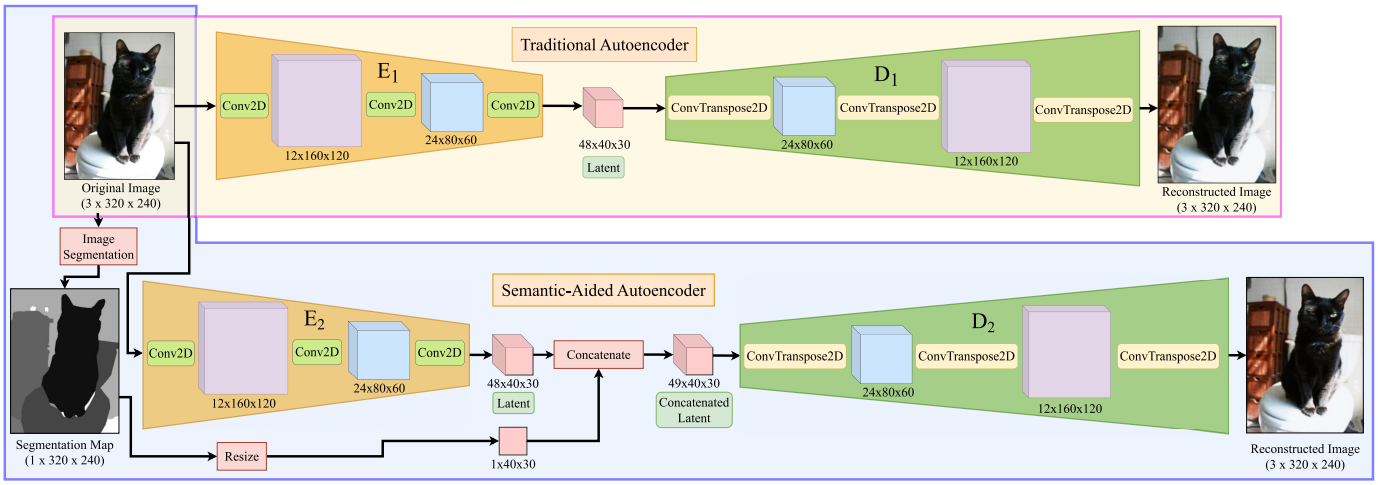


Fig. 1: Architecture of traditional and proposed semantic-aided autoencoders (Training phase)

transmission system and compare its performance with a traditional autoencoder-based system. We demonstrate the advantages of our proposed system in terms of multiple evaluation metrics (mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM)).

- For our proposed semantic-aided system, we use UEP channel coding schemes and compare their performance with equal error protection (EEP) schemes. The UEP schemes are implemented using two methods: (1) By using low-density parity check (LDPC) code for the latent map and a simple Hamming code for the segmentation map; (2) By utilizing the intrinsic UEP property of polar codes.

II. System Model

Our proposed semantic-aided autoencoder-based image transmission system operates in two phases, the training phase and the communication phase.

A. Training Phase

In the training phase, the autoencoder learns to create a latent map from the original image and then reconstruct the image from this latent map. The *encoder* generates the latent map, and the reconstruction is done by the *decoder*. The training phase of the traditional approach is shown in Fig. 1. The autoencoder is implemented as a convolutional neural network. The encoder E_1 consists of three *Conv2D* layers, while the decoder D_1 consists of three *ConvTranspose2D* layers. Fig. 1 also shows the architecture of our proposed semantic-aided autoencoder. The encoder E_2 has the same architecture as the encoder E_1 of the traditional approach. The architecture of the decoder D_2 is different from that of decoder D_1. In our approach, apart from the latent map, the decoder is provided with additional semantic information in the form of the segmentation map of the image. The segmentation map is resized and concatenated as an extra channel to the latent map. The decoder D_2 takes this concatenated latent map and reconstructs the original image using three *ConvTranspose2D* layers. All the layers in the encoders E_1 and E_2 have a ReLU activation function. For decoders $D1$ and $D2$, all layers except the last one, use *ReLU* activation function, while the final layer employs a *sigmoid* activation function to produce the normalized pixel values of the reconstructed image.

B. Communication Phase

The communication phases of the traditional approach and of our semantic-aided approach are shown in Fig. 2. In this phase, the trained *encoder* (E_1 for the traditional approach and E_2 for the semantic-aided approach) is placed at the sender, and the trained *decoder* (D_1 for the traditional approach and D_2 for the semantic-aided approach) is placed at the receiver of the communication system. In our semantic-aided system, the trained encoder-decoder pair serves as the common knowledge base between the transmitter and the receiver.

When transmitting a new image, the sender employs the trained *encoder* to generate the latent map of the image. This latent map undergoes channel coding and modulation. It is then transmitted over a noisy communication channel. At the receiver, a noisy version of the latent map is recovered after channel decoding and demodulation. The decoder uses the noisy version of the latent map to reconstruct the image. Due to noise in the latent data, the quality of the reconstructed image is lower as compared to the original image. In our semantic-aided approach, the difference is that the sender concatenates the generated latent map with the segmentation map of the image. This combined latent and segmentation map then undergoes channel coding and modulation. The decoder at the receiver uses the received concatenated latent map as its input to reconstruct the original image. The segmentation map is extra semantic information that is used for improving the quality of the reconstructed image in our approach. While transmitting an image, the latent map values are encoded using an 8-bit representation while the segmentation map values are encoded using a 4-bit representation. In our experiments, we find that using a lower-resolution representation for the

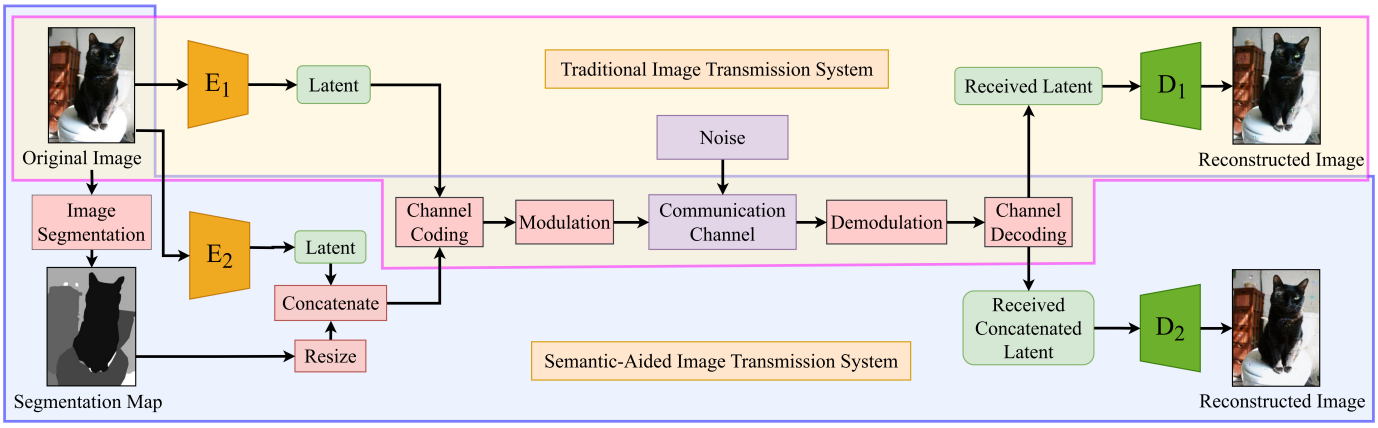


Fig. 2: Communication phase of traditional and proposed semantic-aided autoencoder-based image transmission systems

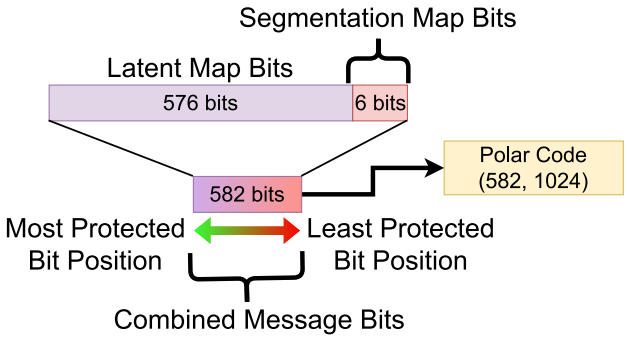


Fig. 3: Polar code-based UEP scheme

semantic data does not lead to a loss in the quality of the reconstructed image. This is because of the intrinsically greater resilience of semantic data to channel noise [1].

III. Equal and Unequal Error Protection

In the communication phase of our semantic-aided image transmission system, both the latent and the segmentation map bits traverse a noisy channel, leading to degradation in the quality of the reconstructed image. To negate the effect of channel noise, we use channel coding. In our experiments, we use Hamming codes, LDPC codes, and Polar codes [7]. LDPC codes and Polar codes are being used in the 5G wireless communication protocol [8]. In wireless communication, the quality of the received image can deteriorate drastically when the signal-to-noise ratio (SNR) drops to low values, necessitating the use of error correction codes. It is to be noted that the significance of the latent map data and that of the semantic segmentation map data are different. The receiver primarily needs the latent map to reconstruct the original image, whereas the semantic segmentation map serves as supplementary information. Additionally, semantic data exhibits greater resilience to channel noise compared to traditional data [1]. So, in this heterogeneous data scenario, we propose UEP channel coding schemes. We give more protection to the latent map values and less protection to the segmentation map values. Error correction codes (ECCs) are represented as (K, N) codes, where K is the number of message bits and N is the number of bits in the codeword after the message is encoded. In our implementation, we use $(32400, 64800)$ LDPC codes, $(1013, 1023)$ Hamming codes, and $(582, 1024)$ polar codes.

EEP: The first EEP scheme is the *Full LDPC* scheme. Here, the latent map bits and the segmentation map bits are separately encoded using LDPC codes. In the polar code-based EEP scheme, called *Polar EEP*, the two different types of data bits are encoded separately using polar codes. The EEP schemes protect both the latent and segmentation map data identically.

UEP: The LDPC code-based UEP scheme is called the *LDPC+Hamming* scheme. In this scheme, the latent map bits are encoded using LDPC codes. For the segmentation map bits, we use a much simpler Hamming code. Using a simpler channel coding scheme leads to lower computation complexity. Also, a $(1013, 1023)$ Hamming code adds less number of parity bits to the message bits, as compared to a $(32400, 64800)$ LDPC code, reducing the total number of bits that need to be transmitted. Polar codes have an inherent UEP property due to the underlying process of channel polarization. Different bits of the polar code codeword suffer from different degrees of error when transmitted through the same channel [9]. We utilize this property to provide differential error protection to the latent map and the segmentation map bits in the *Polar UEP* scheme. As shown in Fig. 3, we combine the latent map bits and the segmentation map bits in the same codeword. The latent map bits are assigned to the more protected bit positions of the codeword, while the segmentation map bits are assigned to the less protected bit positions. This combined codeword is then transmitted over a noisy channel.

IV. Experimental Setup

Dataset: We use the COCO-Stuff dataset [10] for our experiments. This dataset has previously been used to design and evaluate a semantic image transmission system [3]. The dataset contains images of different objects along with their segmentation maps. Both the traditional and the semantic-aided autoencoders are trained on a truncated version of

the dataset containing $10,000$ images. The quality of the reconstructed image in the communication phase is compared over a test dataset of 100 images. Both the autoencoders are trained for 50 epochs, with MSE loss as the optimization criterion. For our experiments, we resize the images to a dimension of 320×240 with 3 channels. We choose the latent map size to be 40×30 with 48 channels.

Setup: The autoencoder models are implemented using Python 3 and the Pytorch (version 2.0.1+cu118) framework. All the experiments for the training and communication phases are done on a machine with an NVIDIA GeForce GTX 1080 Ti GPU, an Intel Xeon E5-1650 v3 CPU (32 GB RAM), and the Windows 10 Pro operating system. The noisy communication channel and the channel coding and modulation schemes are implemented in Matlab (version R2023a). In our experiments, we use the quadrature phase-shift keying (QPSK) modulation scheme. Polar codes are implemented using the Matlab 5G Toolbox, while LDPC codes, Hamming codes, and QPSK modulation are implemented using the Matlab Communication Toolbox.

V. Results

A. Performance Metrics

The performance of the traditional and our proposed semantic-aided approaches with various EEP and UEP channel coding schemes are evaluated using three metrics: MSE, PSNR, and SSIM.

The MSE metric measures the average of the squared pixel-wise differences between the corresponding pixels of two images. It is given by

$$\text{MSE} = \frac{1}{NMC} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \sum_{c=0}^{C-1} [I(i,j,c) - \hat{I}(i,j,c)]^2 \,, \quad (1)$$

where N is the image height, M is the image width, C is the number of colour channels ($C = 3$ for RGB), $I(i,j,c)$ is the pixel value of colour channel c of the original image at position (i,j), and $\hat{I}(i,j,c)$ is the pixel value of colour channel c of the reconstructed image at position (i,j).

The PSNR metric quantifies the signal-to-noise relationship, which affects the accuracy of signal representation, and is closely linked to the MSE metric. It is given by

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \,, \quad (2)$$

where MAX represents the maximum possible value of a pixel. In our experiments, all the pixel values are mapped to a range of $[0,1]$. In general, a higher PSNR value suggests that the reconstructed image possesses superior quality.

The SSIM metric, introduced by Wang et al. [11], seeks to emulate human visual system-based image quality perception. It is given by

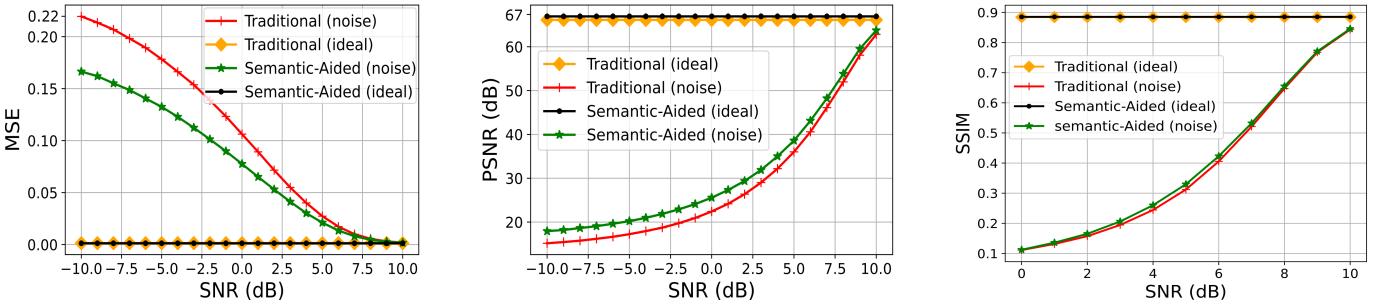
$$\text{SSIM}(I,\hat{I}) = \frac{(2\mu_I\mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \,, \quad (3)$$

where μ_I is the average pixel value of the original image I, $\mu_{\hat{I}}$ is the average pixel value of the reconstructed image $\hat{I}$, σ_I^2 is the variance of pixel values of image I, $\sigma_{\hat{I}}^2$ is the variance of pixel values of image $\hat{I}$, $\sigma_{I\hat{I}}$ is the covariance between pixel values of images I and $\hat{I}$. The SSIM value lies in the range of $[-1,1]$. A value closer to 1 indicates higher structural similarity between the two images, and a value closer to -1 suggests significant dissimilarity. Unlike MSE and PSNR, which rely on absolute error measurements, SSIM provides a comparative evaluation of image quality.

B. Experimental Results

1) Comparison of Traditional and Proposed Semantic-Aided Approaches: We perform experiments for a noisy channel that adds Additive White Gaussian noise (AWGN) to the transmitted values at SNR values in the range of -10 dB to 10 dB. In the first experiment, we compare the quality of the reconstructed image for the traditional and our proposed semantic-aided autoencoders. The comparison is done for the ideal case where we assume that there is no noise in the channel and for a noisy channel at different SNR values. In these experiments, we do not use any channel coding scheme for error correction. The results for this experiment are shown in Fig. 4. In the graph for the MSE metric (Fig. 4a), we see that in the ideal case, labeled as *Traditional (ideal)* and *Semantic-Aided (ideal)*, the MSE between the original image and the reconstructed image is very close to 0. This shows that when the decoder receives a latent map with no errors, it is able to generate a good reconstruction of the original image for both the traditional and our proposed semantic-aided approaches. The difference in performance between these two approaches becomes evident when we transmit over a noisy channel. The MSE in the presence of noise is shown using *Traditional (noise)* and *Semantic-Aided (noise)*. Transmitting the segmentation map of the image in our proposed approach improves the MSE of the received image by up to 23% with only a 2% increase in the bandwidth compared to the traditional approach. The gap in performance is greater at lower SNR values, and the MSE values converge as the SNR improves. For the PSNR graph (Fig. 4b), we observe a similar improvement for the semantic-aided approach with up to 18% improvement at low SNR values. Moreover, in the ideal case, the PSNR for the semantic-aided autoencoder (66.9 dB) is better than that of the traditional one (66.13 dB). In Fig. 4c we see that the SSIM for the semantic-aided approach is better than that for the traditional approach. For SNR below 0 dB, the SSIM value is less than 0 and there is no visible similarity between the original and the reconstructed image. Hence, these values are not plotted in the SSIM graph.

We visually compare the quality of the reconstructed images for the ideal case and with noise at different SNR values. The original and the reconstructed images for the traditional and the proposed semantic-aided approaches are shown in Fig. 5. It is clearly evident from the figure that the semantic-aided approach is able to generate a better reconstruction of the original image compared to the traditional approach. We see that even at SNR $= 1$ dB, the shape of the airplane is already slightly visible in the semantic-aided reconstruction, but in the



(a) Mean square error (MSE) (b) Peak signal-to-noise ratio (PSNR) (c) Structural similarity index measure (SSIM)

Fig. 4: Comparison of traditional and proposed semantic-aided image transmission systems

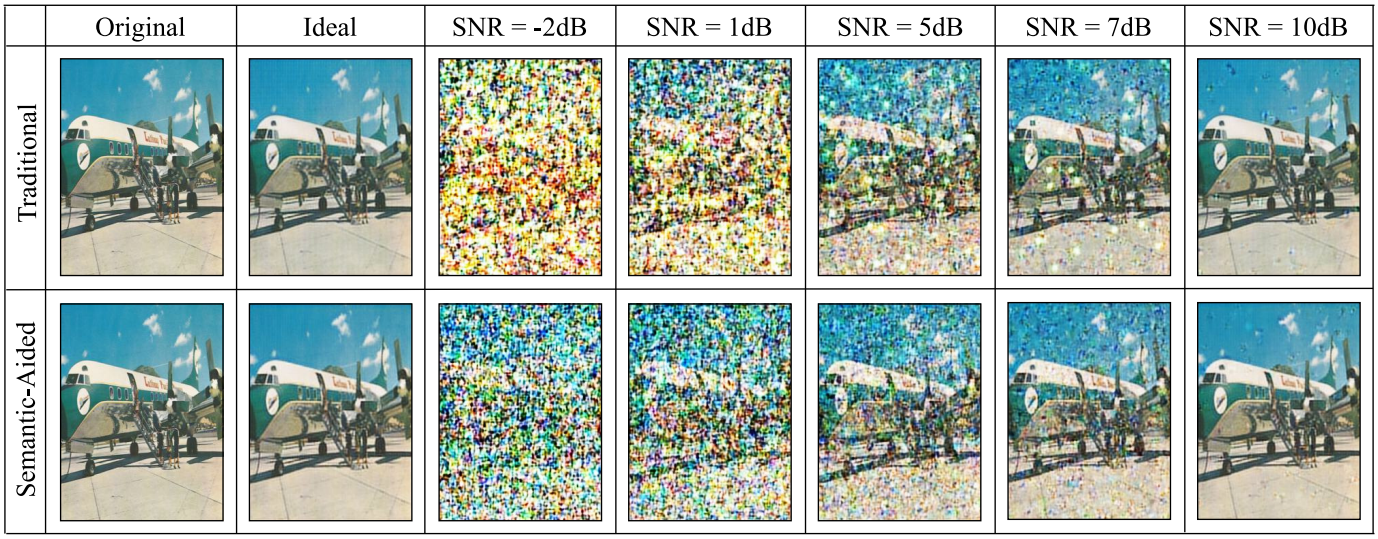
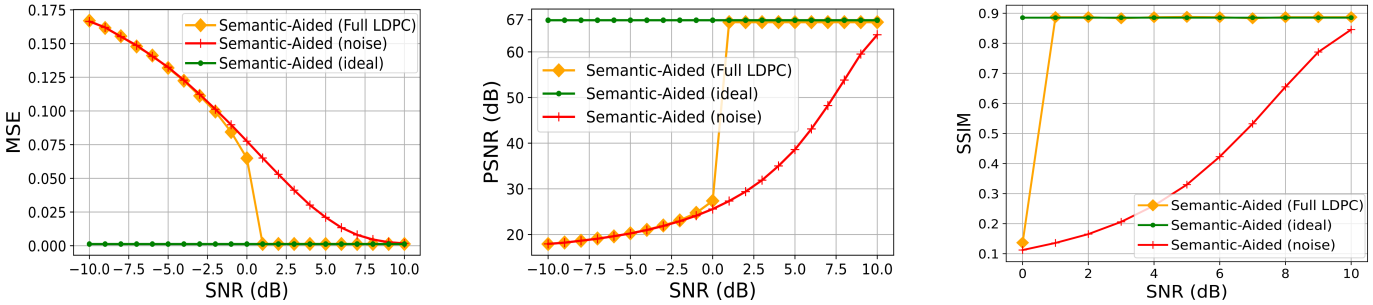


Fig. 5: Visualization of images reconstructed by traditional and proposed semantic-aided approaches

traditional reconstruction, it is not visible at all. Similarly, at SNR $= 5$ dB and SNR $= 7$ dB, we have a much clearer reconstructed image using the semantic-aided approach. This happens because of the extra semantic information provided to the decoder in the form of the segmentation map of the image. It is important to note that this improvement is obtained with an equal number of training epochs for both the autoencoders.

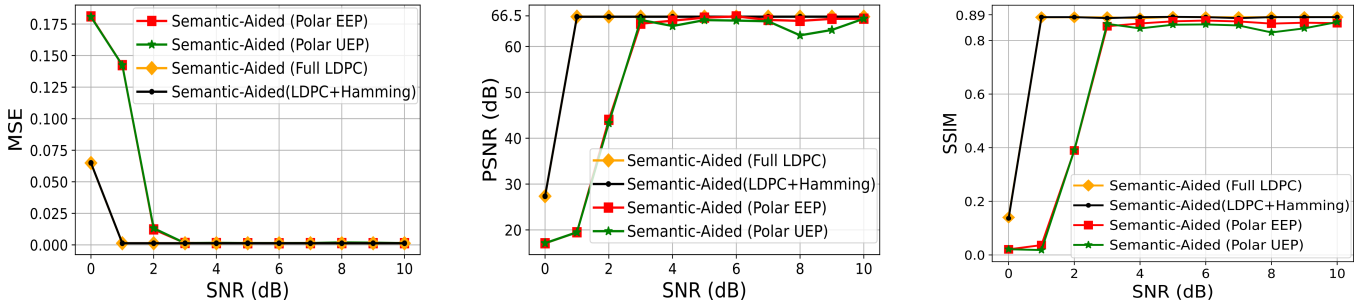
2) Effect of Adding LDPC Channel Coding to the Semantic-Aided Approach: In the second experiment, we add LDPC channel coding to the semantic-aided approach when transmitting over a noisy channel. The results of this experiment are shown in Fig. 6. In these graphs, we compare the quality of the reconstructed image for three different scenarios: (1) ideal case (no noise), (2) noisy channel with AWGN noise but no channel coding, and (3) noisy channel with LDPC channel coding for both the latent and the segmentation map bits. We see in Fig. 6a that when we add LDPC channel coding, the MSE improves sharply compared to the scenario without LDPC coding in the presence of channel noise. At SNR ≥ 1 dB, the quality of the reconstructed image improves and becomes equal to the image quality in the ideal case. A similar trend is observed for PSNR and SSIM, as seen in Fig. 6b and Fig. 6c, respectively. These results show the effectiveness of channel coding to negate the adverse effects of a noisy communication channel.

3) Comparison of EEP and UEP Channel Coding Schemes: In the third experiment, we compare the performance of EEP and UEP channel coding schemes. These experiments are done for a noisy channel with AWGN noise having SNR in the range of 0 dB to 10 dB. The results of these experiments are shown in Fig. 7. In the result graphs, we see that the MSE, PSNR, and SSIM for *Full LDPC* coding and *LDPC+Hamming* coding are the same. The graphs overlap with each other. This shows that even when we provide less error protection to the segmentation map bits using a simple Hamming code, there is no degradation in the quality of the reconstructed image. This observation allows us to replace LDPC codes with much simpler Hamming codes for the segmentation map bits and reduces the computational complexity of channel coding. The result graphs also compare the performance of the *Polar EEP* and the *Polar UEP* schemes. The MSE graph in Fig. 7a shows that there is no degradation in the MSE of the reconstructed image in the *Polar UEP* scheme as compared to the *Polar EEP* scheme. The PSNR and SSIM values are also very close to each other over the entire SNR range. There is no degradation in the quality of the received image. An important observation from the three graphs in Fig. 7 is that LDPC codes have a



(a) Mean square error (MSE) (b) Peak signal-to-noise ratio (PSNR) (c) Structural similarity index measure (SSIM)

Fig. 6: Effect of adding LDPC channel coding to the proposed semantic-aided image transmission system



(a) Mean square error (MSE) (b) Peak signal-to-noise ratio (PSNR) (c) Structual Similarity index measure (SSIM)

Fig. 7: Comparison of equal and unequal error protection channel coding schemes

better error correction capability than polar codes and lead to a more faithful reconstruction of the image at the receiver.

VI. Conclusion

In this paper, we propose a semantic-aided autoencoder-based image transmission system that has up to 23% improvement (in terms of MSE) and 18% improvement (in terms of PSNR) in the quality of the received image with only 2% extra bandwidth requirement over a traditional autoencoder-based system. Furthermore, through a comparative analysis of EEP and UEP channel coding schemes for the semantic and syntactic data, we find that the UEP schemes offer greater efficiency without any degradation in the quality of the received image. Such semantic-aided architectures are stepping stones in the transition from traditional to purely SC systems.

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, Feb. 2022.

[2] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, Jan. 2023.

[3] M. U. Lokumarambage, V. S. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," *IEEE Access*, vol. 11, pp. 37 149–37 163, Apr. 2023.

[4] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, May 2019.

[5] A. Li, X. Liu, G. Wang, and P. Zhang, "Domain knowledge driven semantic communication for image transmission over wireless channels," *IEEE Wireless Communications Letters*, vol. 12, no. 1, pp. 55–59, Jan. 2023.

[6] S. Borade, B. Nakiboğlu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5511–5539, Dec. 2009.

[7] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[8] B. Tahir, S. Schwarz, and M. Rupp, "BER comparison between convolutional, turbo, ldpc, and polar codes," in *Proc. IEEE 24th International Conference on Telecommunications (ICT)*, Limassol, Cyprus, Aug. 2017, pp. 1–7.

[9] A. Shreshtha, P. Singla, and S. R. Sarangi, "Polar code-based approximate communication system for multimedia web pages," in *Proc. IEEE 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, Milan, Italy, May 2022, pp. 537–538.

[10] H. Caesar, J. Uijlings, and V. Ferrari, "Coco-stuff: Thing and stuff classes in context," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, USA, Jun. 2018, pp. 1209–1218.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.