Image transmission using Semantic Communication for Home Intrusion System

K R Nandakishore[1], Rayani Venkat Sai Rithvik [2]

***Abstract*— Semantic Communication, a paradigm where only the necessary information is transmitted, discarding the rest, leading to huge gain in number of bits used and hence saving energy. In the context of home intrusion system, this paper aims at Image transmission using Semantic Communication, when an alarm is triggered.**

I. Introduction

Home Intrusion Systems are a common practice nowadays, to ensure protection against any alarm triggering anomalies. These include sophisticated networks that communicate wirelessly as a part of the IoT environment. Many optimizations can be made to ensure these communications are more environmental friendly. Through Semantic Communication Paradigm, a great reduction in number of bits send is possible. Sending an image itself, would require 8 bits per pixel. Clearly, sending a only the required part, as is meant by Semantic Communication, will requried lesser number of bits. This paper explores the implementation of Semantic Communication paradigm within this framework and observes the immense gain we obtain.

II. Literature Survey

In [1], M. U. Lokumarambage et al. implement an end-to-end image transmission system using semantic communication. In the pipeline, they extract the semantic map from the input image and encode it using Polar Codes. The semantic map is sent through the Additive White Gaussian Noise(AWGN) channel with the Binary Phase Shift Keying(BPSK) modulation. At the receiver, the demodulation and decoding of the semantic map take place. Using the shared knowledge base between the transmitter and receiver, a Generative Adversarial Network(GAN) is used to reconstruct the image based on the semantic map. Error concealment is done by applying the median filter on the output. It is to be noted that the COCO-Stuff dataset is used in the paper. The Peak Signal-to-Noise Ratio(PSNR) is used to measure the effect of varying noise on the semantic maps. As the GAN is trained on a particular dataset, it is to be noted that the images generated from the semantic map is not going to follow the exact details of the transmitted image. The generation depends on the shared knowledge database and hence cannot regenerate the same image.

In [2], the focus is more on the effect of varying the polar code's block length and rate on the Signal-to-Noise Ratio(SNR). It proposes a flexible simulation software that transmits semantic segmentation maps over a communication channel. Bit Error Rate(BER) and Frame Error Rate(FER) are the primary metrics in the paper.

[1]EE21BTECH11027, Department of Electrical Engineering, Indian Institute of Technology Hyderabad, India.

[1]EE21BTECH11043, Department of Electrical Engineering, Indian Institute of Technology Hyderabad, India.

III. Implementation Pipeline

A. Pipeline Overview

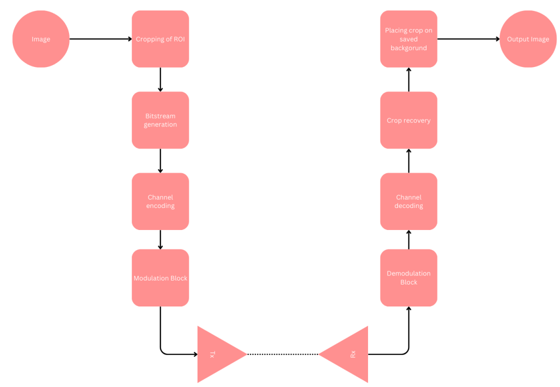


Fig. 1: Pipeline overview

As infered from fig.1, a crop of the region of interest (RoI) is obtained from the image. This is done using semantic segmentation, using fastseg library in python. The obtained crop of RoI is converted to a bitstream, channel encoded, modulated and then transmitted. At the receiver side, the bitstream is parsed appropriately and the crop of RoI is recovered. It is to be noted that, in this specific application, the background remains same and hence is a common knowledge between the transmitter and receiver. The crop of RoI is then placed at the exact location in the background, hence recreating the original image. In this paper, we consider people as RoI.

B. Cropping of RoI

fastseg is a semantic segmentation library available in Python that is used here to obtain the semantic map. The RoI is obtained by taking only the pixel values with class labels of interest. The minimum and maximum x and y coordinates of such pixels are found to create a bounding box, which is then considered the RoI crop. Within the bounding box, any pixel of not interest, will be flipped to zero. This can be observed in fig.2.

C. Bitstream Generation

In order to properly reconstruct the RoI crop as well as place it in the background, along with the crop, we need to send the coordinates of the bounding block as well as the width and height of it as well.

A basic assumption taken is that the CCTV footage is going to be of resolution 512x512. This will ensure that we will need only 9 bits to represent the coordinates, width and height of the bounding boxes. In the RoI crop, each pixel will need 8 bits.

The parsing scheme followed is the first 36 bits represent the coordinates, width and height in the fashion [x, y, w, h]. The rest of the bits represent the flattened RoI crop. This is then channel encoded and transmitted.

D. Channel encoding and Modulation

Polar codes are used to encode the bitstream. Let the block length be N. The polarization phenomenon divides the physical channel into reliable(K bit positions) and unreliable(N-K bit positions) virtual channels. The code rate is $\frac{K}{N}$. The K most reliable bit positions are included in the information set, and the remaining N-K bit positions are included in the frozen set. As the block length increases, the reliabily of the K channels increases. The encoded symbols are then mapped using the BPSK modulation scheme and transmitted over AWGN channel, after which the demodulation and decoding is done. The entire channel encoding and decoding, modulation and demodulation, and transmission is simulated using the Python library polarcodes

E. Recovery of RoI Crop

After the bitstream is received, we can obtain the coordinates, width and height from the first 36 bits. The rest of the bits are also converted back to uint8 format to obtain each pixel. Thus we obtain a flattened vector, which is resized to (height, width) to reconstruct the RoI crop. The reconstructed RoI crop can be observed in (c), Fig.2.

F. Placing RoI crop on background

Background is a fixed entity throughout this process, making it needless to transmit. The RoI crop reconstructed is placed at the coordinates obtained so as to reconstruct the original image intended to be transmitted.

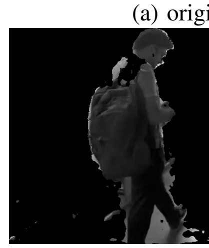
The RoI crop is placed such that wherever the pixel value is not zero in the RoI crop, we place that pixel in the background. This results in an almost similar image compared to the one originally intended to be transferred. This is seen in (d), Fig.2

IV. Observations and Results

We can observe the outputs from the entire pipline in fig.2. This figure ensures the quality of technique followed and shows how promising the Semantic Communication paradigm can be. The images were transmitted at an SNR of 5 dB and a block length of 64, at a rate of 1/2.

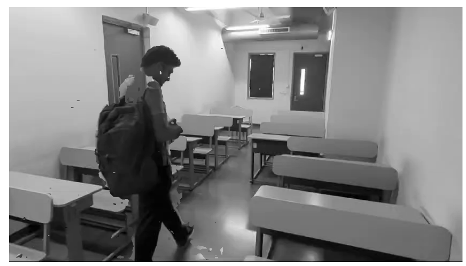
In the table I, the received image and original image are compared using standard image comparison metrics. The

(a) original image



(b) RoI crop

(c) RoI crop reconstructed



(d) RoI crop placed on background

Fig. 2

Metric	value
MSE	1956.015
PSNR	15.217
MSSIM	0.649

TABLE I: Comparing original image and received image

metrics reveal how good the received image is compared to the original image.

The main goal here is the gain obtained from saving bits transmitted. The sample image shown in (a), fig2, requires 3242770 bits to be transmitted for proper reconstruction at receiver. Whereas following the Semantic Communication paradigm, we reduce the number of bits transmitted to 1253916 bits, saving 1988854 bits, effectively saving is 61.332%.

From the BER vs SNR plot, we see that for lower SNRs, the polar codes with less block length have lesser BER than the codes with higher block length. For SNRs more than 3, the polar codes with higher block length have lesser BER than the codes with more block length. This is because of higher polarization of larger polar codes which means that some channels' reliability increases with increase in block length.

V. Conclusion

Through the exploration in this paper, we can gaurantee the effectiveness of Semantic Communication paradigm in

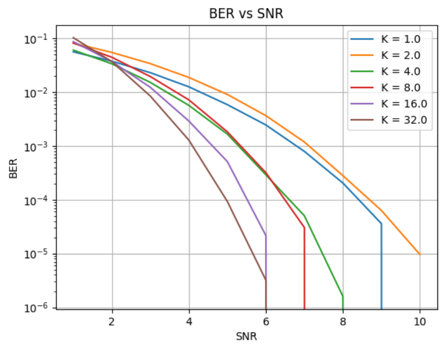


Fig. 3: BER vs SNR plot with rate 0.5

terms of reduction of bits transmitted, resulting in energy savings. As the received image is similar to the original image, this paradigm proves itself to be of practical use with an acceptable level of quality.

VI. Future Work

As could be observed the observations, the RoI crop is not exactly acccurate. By making use of traditional image processing techniques like minor blob removal and dilation, we can get a closed and accurate RoI.

A detailed system for detecting anomalous events could be designed and integrated along with the camera, which could trigger an alarm making the further resultant steps more timely.

VII. Codes

The project repository can be found here.

VIII. Acknowledgement

We thank Dr.Zafar Ali Khan and the course TAs for their support and guidance throughout and the course and the project.

References

[1] 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. 37149–37163, 2023.

[2] H. Rezaei, T. Sivalingam, and N. Rajatheva, "Automatic and flexible transmission of semantic map images using polar codes for end-to-end semantic-based communication systems," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–6, 2023.