Image transmission using Semantic Communication for Home Intrusion System

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

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

I. Introduction

Home Intrusion Systems is 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 environmentally friendly. Through the Semantic Communication Paradigm, a great reduction in the number of bits sent is possible. Sending an image itself would require 8 bits per pixel. Clearly, sending only the required part, as is meant by Semantic Communication, will require a lesser number of bits. This paper explores the implementation of the Semantic Communication paradigm within this framework and observes the immense gain we obtain.

II. Literature Survey

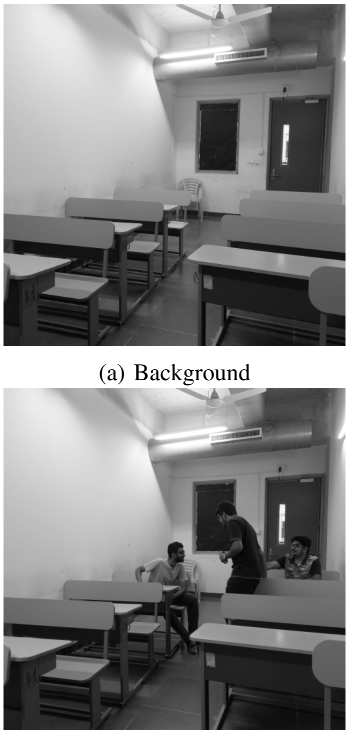
In [1], M. U. Lokumarambage et al. implement an end-to-end image transmission system using semantic communication. In the pipeline, they extract the semantic map from the input image and encode it using Polar Codes. The semantic map is sent through the Additive White Gaussian Noise(AWGN) channel with 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.

III. Dataset

The dataset used for the project is a custom one we developed. There are 54 images in the dataset, with 3 background images(which is the shared knowledge between the transmitter and receiver) included. Each background image has 18 images associated with it, with six images containing a single person, six consisting of two people, and the other six consisting of three people. The dataset was collected through a mobile phone, so there might have been small inconsistencies between the background and the respective images.

The dataset can be accessed here.



(a) Background

(b) Background with people

Fig. 1: Dataset samples

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

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

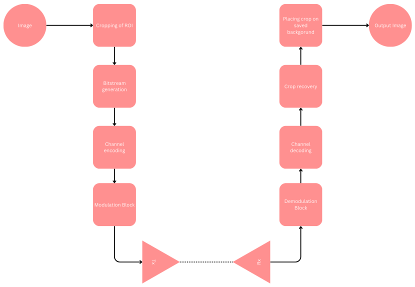


Fig. 2: Pipeline overview

IV. Implementation Pipeline

A. Pipeline Overview

As inferred from fig.2, a crop of the region of interest (RoI) is obtained from the image. The obtained crop of RoI is converted to a bitstream, which is 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 constant and, hence, is the shared 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 the class ”people” as RoI.

B. Cropping of RoI

Two methods are explored to get the RoI crop:

1) Using Semantic Segmentation: Semantic Segmentation is a visual task that is used to classify each pixel in an image. 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. Here, the maximum and minimum values of the x and y coordinates of the pixels in RoI are used to obtain the bounding box coordinates.

2) Using Object Detection: Object Detection is a visual task that detects certain objects within an image and gives the coordinates to a bounding box of that class. Here, a lightweight implementation of YoloV3 is used. This GitHub Repository was used for object detection.

We get the RoI crop using the bounding box coordinates from either method.

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 that 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 reliability of the K channels increases. The encoded symbols are then mapped using the BPSK modulation scheme and transmitted over the AWGN channel, after which the demodulation and decoding are done. The entire channel encoding and decoding, modulation and demodulation, and transmission are 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 the ’uint8’ format to obtain each pixel. Thus, we obtain a flattened vector, which is then resized to the tuple (height, width) to reconstruct the RoI crop. The reconstructed RoI crop can be observed in (c), Fig.5.

F. Placing RoI crop on background

The 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 exactly at the position mentioned by the coordinates obtained. This results in an almost similar image compared to the one originally intended to be transferred. This is seen in (d), Fig.5

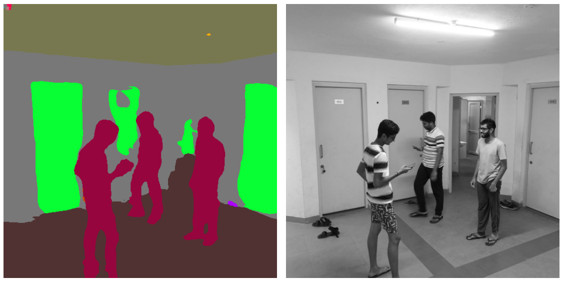
G. Fallback option

There could be cases where the model fails to detect the RoI, which enforces the need to have a fallback option. In such cases, we choose to send the whole image itself.

V. Metrics

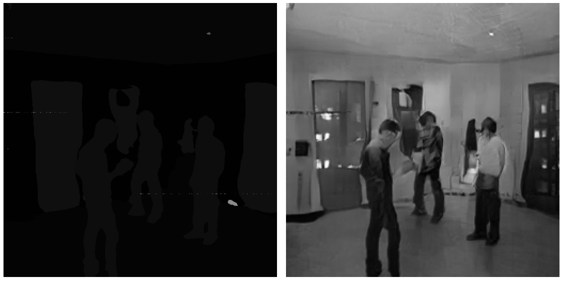
To compare the resultant reconstruction and the ground truth images, the following metrics are used:

- Normalized version of the MSE loss.
  NMSE loss = $\frac{1}{NM}\sum_i\sum_j \frac{[I(i,j)-\hat{I}(i,j)]^2}{I(i,j)^2}$
- PSNR (Peak Signal to Noise Ratio) = $10\log_{10}\frac{255^2}{\text{MSE Loss}}$
- SSIM (Structural Similarity index) (-1 $\leq$ SSIM $\leq$ 1). This metric measures how similar two images are.



(a) semantic map (colored) (b) Ground Truth

Fig. 3: Baseline result sample



(a) semantic map received (b) Reconstructed Image

Fig. 4: Baseline result sample

Baseline results			
NMSE	**PSNR**	**MSSIM**	**bits saved (%)**
21.196	11.084	0.433	91.574%

TABLE I: Baseline Statistics

VI. Baseline results

For observing the baseline results, Segformer with MiT-B3 backbone was used for semantic segmentation and Spade model was used for image generation.

The results can be observed in fig.4 and Tab.I. The bits saved here were calculated on the basis of ground truth file size and semantic map size.

VII. Observations and Results

As discussed earlier, two methods are followed to get the RoI crop: Semantic Segmentation and Object Detection. The following describes the observations and results.

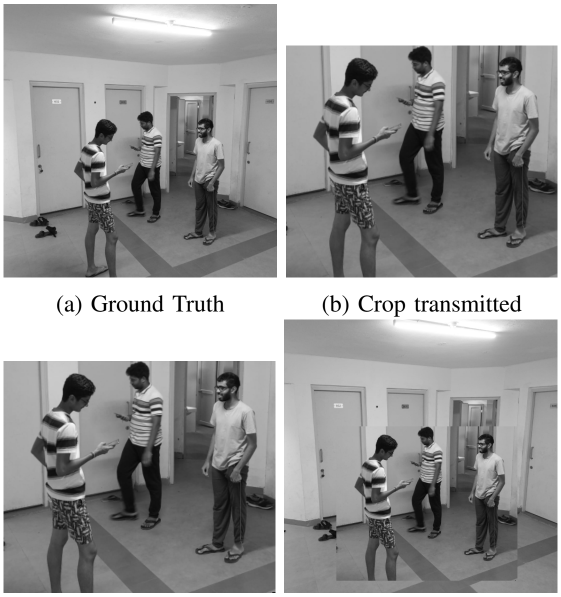
A. Using Semantic Segmentation

The results for a sample from the dataset can be observed from fig.5.

NOTE: A small displacement can be seen in the reconstruction. This is due to small inconsistencies present in the dataset itself. However, the overall goal of transmitting only the necessary information can be visualized here.

Fig. 6 shows the BER vs SNR plots for different K values in semantic segmentation.

The table II shows the average statistics over the entire dataset.



(a) Ground Truth (b) Crop transmitted

(c) Crop received (d) Reconstructed image

Fig. 5: Results using Semantic segmentation

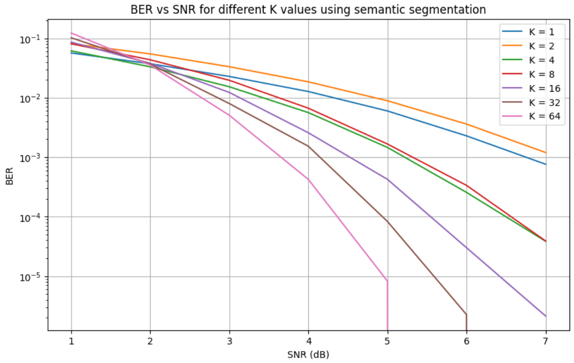


Fig. 6: BER vs SNR for semantic segmentation

B. Using Object Detection

The results for a sample from the dataset can be observed from fig.7.

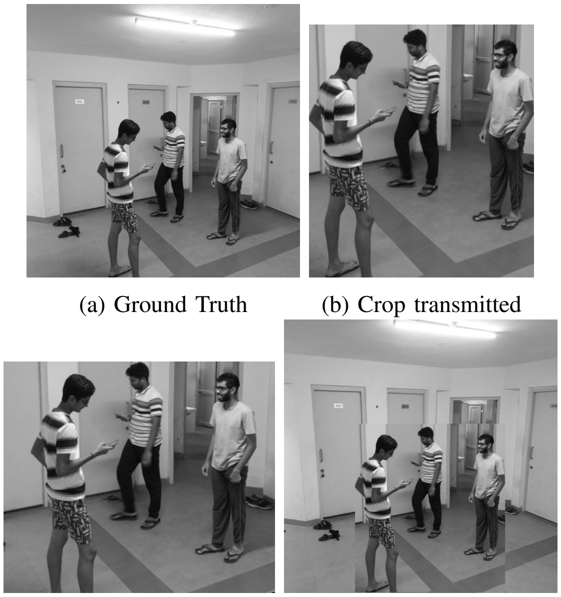
Fig. 8 shows the BER vs SNR plots for different K values in object detection. For both object detection and semantic segmentation, we see that for lower SNRs, the polar codes with less block length have lesser BER than those with higher block length. For SNRs over 2, the polar codes with higher block lengths have lesser BER than those with more block lengths. This is because of the higher polarization of larger polar codes, which means that some channels' reliability increases with an increase in block length.

C. Comparative Remarks

Based on Table II, using the image comparison metrics, we can see that the semantic segmentation method outperformed the object detection method. However, we can observe that the percentage of bits saved is much higher for the Object Detection Method than for the Semantic Segmentation Method. This is due to the following issue, which

Method to get RoI	NMSE	PSNR	MSSIM	bits saved (%)
Semantic Segmentation	3.912	36.981	0.784	54.415%
Object Detection	5.412	18.377	0.632	86.718%

TABLE II: Average Statistics



(a) Ground Truth (b) Crop transmitted

(c) Crop received (d) Reconstructed image

Fig. 7: Results using Object Detection

was resolved by the fallback option. Using the Semantic Segmentation Method, in many images within the dataset, it fails to identify the RoI and hence sends the full image itself, boosting the reconstruction metric results, whereas, for the Object Detection Method, we get up to 95% average reduction in the case when a single person is present, over all the backgrounds in the dataset. These metrics depend on how well the model performs.

These results show that the Object Detection Method is more effective and suited for the task at hand.

VIII. Complexity Analysis

The image 2.png from person 3 in background 3 is used for complexity analysis.

For the Semantic Segmentation Method, for the DL model a total of 1.7345e+10 operations and 3284547 parameters are present and for channel encoding a total of 36910469 operations is done.

For the Object Detection Method, for the DL model a total of 9.9829e+10 operations and 61949152 parameters are present and for channel encoding a total of 49491809 operations is done.

For calculating the number of operations and parameters of the DL models, pytorch-toolbox's profiler was made use of.

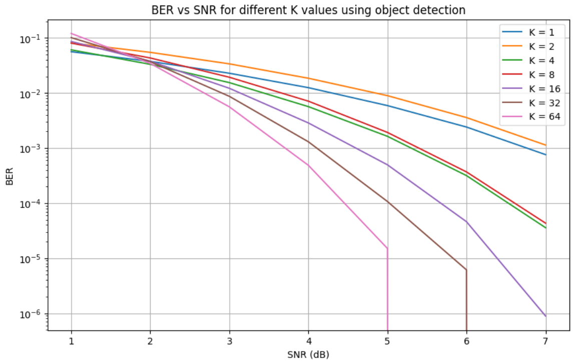
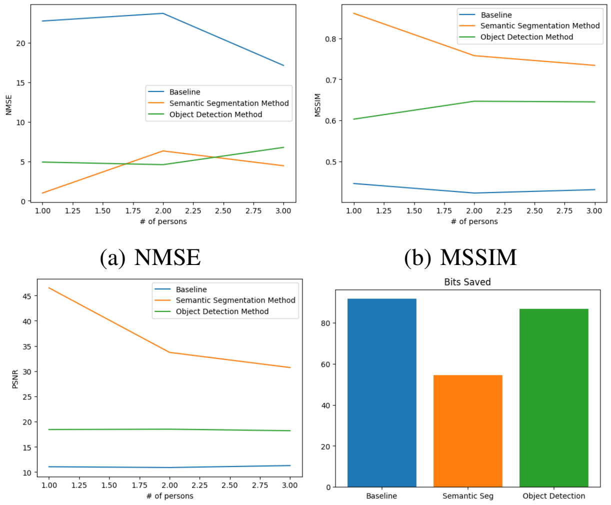


Fig. 8: BER vs SNR for object detection



(a) NMSE (b) MSSIM

(c) PSNR (d) Bits Saved (Average over dataset)

Fig. 9

IX. Conclusion

Through the exploration in this paper, we can guarantee the effectiveness of the Semantic Communication paradigm in terms of the reduction of bits transmitted, resulting in energy savings. As the received image is similar to the original image, this paradigm proves to be of practical use with an acceptable level of quality.

This paper explores two methods to find RoI: the Semantic Segmentation Method and the Object Detection Method. As seen from observations, the task is better achieved using the Object Detection Method, as the Segmentation Method sometimes fails to detect the RoI. It could also be observed that the RoI detected using the Semantic Segmentation Method can also be much smaller than the actual RoI, making the Object Detection Method a better choice.

X. Possible Improvements

For the Object Detection Method, multiple bounding boxes are obtained from the model, from which a universal bounding box is derived. Another path is to send these individual bounding boxes, thereby reducing the number of bits transmitted by a slightly higher margin. This can be more effective if two classes belonging to RoI are present at opposite ends of the image, where the universal bounding box will include the in-between unnecessary area, whereas if we send it as two blocks, then the number of bits transmitted will be significantly reduced.

XI. Codes

The project repository can be found here.

XII. Referred and Used codes

- fastseg
- YoloV3 lightweight implementation
- polarcodes python implementation
- Semantic segmentation repo
- Spade GAN
- pytorch-toolbox

XIII. 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.