Semantic segmentation-based semantic communication system for image transmission

Jiale Wu, Celimuge Wu, Yangfei Lin, Tsutomu Yoshinaga, Lei Zhong, Xianfu Chen, Yusheng Ji

PII: S2352-8648(23)00041-X

DOI: https://doi.org/10.1016/j.dcan.2023.02.006

Reference: DCAN 625

To appear in: *Digital Communications and Networks*

Received Date: 31 October 2022

Revised Date: 8 January 2023

Accepted Date: 10 February 2023

Digital Communications and Networks(DCN)

HOSTED BY ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/dcan

Semantic Segmentation-based Semantic Communication System for Image Transmission

Jiale Wu[a], Celimuge Wu*[a], Yangfei Lin[a], Tsutomu Yoshinaga[a]
Lei Zhong[b], Xianfu Chen[c], Yusheng Ji[d]

[a]The University of Electro-Communication, Tokyo 182-8585, Japan
[b]Toyota Motor Corporation, Tokyo 112-8701, Japan
[c]VTT Technical Research Centre of Finland, Oulu, Finland
[d]National Institute of Informatics (NII), Tokyo 101-8430, Japan

Abstract

With the rapid development of artificial intelligence and the widespread use of the Internet of Things, semantic communication, as an emerging communication paradigm, has been attracting great interest. Taking image transmission as an example, from the semantic communication's view, not all pixels in the images are equally important for certain receivers. The existing semantic communication systems directly perform semantic encoding and decoding on the whole image, in which the region of interest cannot be identified. In this paper, we propose a novel semantic communication system for image transmission that can distinguish between regions of interest (ROI) and regions of non-interest (RONI) based on semantic segmentation, where a semantic segmentation algorithm is used to classify each pixel of the image and distinguish ROI and RONI. The system also enables high-quality transmission of ROI with lower communication overheads by transmissions through different semantic communication networks with different bandwidth requirements. An improved metric $\theta PSNR$ is proposed to evaluate the transmission accuracy of the novel semantic transmission network. Experimental results show that our proposed system achieves a significant performance improvement compared with existing approaches, namely, existing semantic communication approaches and the conventional approach without semantics.

KEYWORDS:
semantic communication, semantic segmentation, image transmission, image compression, deep learning

1. Introduction

With the advent of the information age, information has gradually become a part of people's lives [1]. However, the traditional grammatical information compression and transmission are approaching the limit of Shannon's information theory [2], and the current capacity and other performance improvement methods cannot support the continuous development of communications in the future [3]. As early as 1948, Shannon mentioned three levels of information dissemination in his paper [4], technical problems, semantic problems and effect problems. However, subject to the level of technology at that time, it was impossible to solve the problem of semantics and effect well. With the recent rise of artificial intelligence, natural language processing and computer vision technologies continue to improve, and people can extract semantics from information, which provides the possibility for the next generation of communication systems [5].

Compared with traditional communication, semantic communication has the advantage of lower latency, less bandwidth, and higher throughput [6]. An extraordinary feature of semantic communication is its

*Corresponding author
[1]E-mail address: g2241007@edu.cc.uec.ac.jp (J. Wu); celimuge@uec.ac.jp (C. Wu); linyangfei@uec.ac.jp (Y. Lin); yoshinaga@uec.ac.jp (T. Yoshinaga); lei-zhong@toyota-tokyo.tech (L. Zhong); xianfu.chen@vtt.fi (X. Chen); kei@nii.ac.jp (Y. Ji);

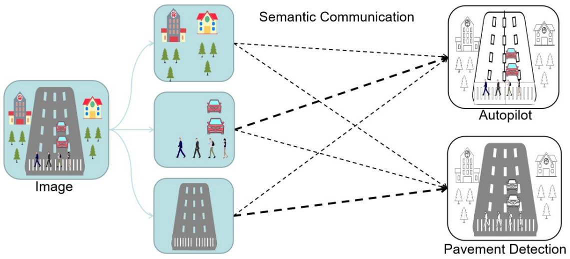


Fig. 1. Different tasks focus on different information.

robustness [7]. The data transmitted by traditional communication is a bit stream, and noise is introduced in the transmission process, which leads to information distortion. While semantic communication transmits features, subtle changes in features will not affect the understanding of the receiver. Therefore, the accuracy of semantic communication is usually higher than traditional communication in a noisy environment.

For image transmission, different tasks focus on different information (the interests of receivers could be different for different transmissions). As shown in Fig. 1, taking an example of a street image, the autopilot application is more interested in understanding the situation of vehicles and people on the road, while the pavement detection application is more concerned with information about road damage. At the same time, we believe that although regions of non-interest (RONI) is not as important as regions of interest (ROI) per task, it should be retained as it still includes a certain amount of information.

Semantic segmentation is one of the key problems in computer vision nowadays [8]. Semantic segmentation is the task of clustering ring parts of an image that belong to the same object class. It is a form of pixel-level prediction because each pixel of an image is classified according to a category. Images can be divided into ROI and RONI according to different tasks using semantic segmentation [9].

Although semantic communication has many advantages, it still challenges bandwidth limitations. The larger the bandwidth, the better the transmission effect. However, when the network bandwidth is limited, the conventional semantic communication approach cannot satisfy the application requirements since all the pixels, including important parts and non-important parts, are treated equally, resulting in losses of some important information. To ensure transmission efficiency under a limited bandwidth, the ROI can be transmitted with more details, and the RONI can be transmitted with fewer details.

This paper proposes a semantic segmentation-based semantic communication system for image transmission. First, according to the requirement of the task, the image is divided into ROI and RONI by using semantic segmentation technology. Then the semantic communication transmission is carried out through different semantic channels. The images obtained after transmission are combined and restored to the original images at the receiver side. The main contributions of this paper can be summarized in four parts:

- We propose a semantic segmentation-based semantic communication system for image transmission that can effectively find ROI from images and transmit it more accurately.
- A data compression algorithm is proposed to reduce communication overhead. Experimental results show that the communication overhead after data compression is only about a quarter of that without data compression.
- For the performance measurement of transmission accuracy of ROI, we propose an improved $\theta PSNR$ metrics based on PSNR.
- We conduct exhaustive evaluations to show that the proposed system outperforms traditional communication systems and single semantic channel communication approach, especially in improving the robustness of the system under low signal-to-noise (SNR) ratio conditions.

The rest of this paper is organized as follows. A brief review of semantic communication related work is given in Section 2. In Section 3, the system model is proposed, and the details of the system model are described. Section 4 shows the performance of the system model on specific tasks. Finally, conclusions are drawn in Section 5.

2. Related works

In this section, we first introduce the origin and development of semantic communication. Then, state-of-the-art technologies about semantic communication and image transmission are discussed.

2.1. Semantic communication

Weaver and Shannon first propose the concept of semantic communication [4]. It includes three levels of communication: the first level is technical issues, which mainly solve the problem of “how to transmit communication symbols accurately", and the second level is semantic issues, which mainly solves the problem of “how does the transmitted symbol accurately convey the meaning"; the third level is the effect problem, which mainly solves the problem of “how the received meaning effectively affects the behavior in the desired way".

For a long time, communication has remained at the first level of solving technical problems. But there are still some scholars trying to solve the semantic problem. Carnap and Bar-Hillel used probabilistic logic theory to study how to measure the amount of semantic information contained in a sentence [10]. They believed that semantic information can be defined by the logical probability of sentence content. Floridi found a fundamental paradox in the theory of semantic communication established by Carnap and Bar-Hillel, specifically, any fact that contradicts with itself will have an infinite amount of information. To solve this problem, Floridi proposed that the amount of semantic information should be represented by the distance from the real event [11]. Bao summarized previous work on semantic communication and proposed a model-theoretic approach to semantic data compression and reliable semantic communication [12]. This approach is linked to Shannon's statistical measure of information and shows that Shannon's source coding theorem and channel coding theorem have semantic counterparts. These pioneering works are all based on logic, but there is a lack of general mathematical models that can express semantic information, so semantic communication has been in its infancy for a long time.

Benefiting from the improvement of hardware and the continuous development of artificial intelligence algorithms, traditional communication has been rapidly transformed into semantic communication. It is inevitable to shift traditional communication to semantic communication.

Xie et al. proposed a deep learning-based semantic communication system [5], called DeepSC, for text transmission. Based on transformers, DeepSC aims to maximize system capacity and minimize semantic errors by restoring the meaning of sentences. Weng et al. designed a deep learning (DL)-based speech signal semantic communication system named DeepSC-S [13]. To improve the restoration accuracy of speech signals, especially the restoration accuracy of basic information, DeepSC-S is developed based on the attention mechanism using the squeeze and excitation (SE) network. Xie et al.'s deep learning-based multi-user semantic communication system is used to transmit unimodal data and multimodal data [14], and a unique transformer-based framework is proposed to unify different tasks. Xie et al. also proposed a DL-based lean distributed semantic communication system (L-DeepSC) for low-complexity text transmission [15]. L-DeepSC can reduce the bandwidth required for communication by pruning model redundancy and reducing weight resolution. However, these semantic communication models do not consider the receiver's demand for information, but transmit all the information, which will lead to an waste in communication overhead.

2.2. Image transmission

Many people have done related work on image transmission. Bourtsolatze proposed a joint source and channel coding technique for wireless image transmission [16], which directly maps image pixel values to actual complex-valued channel input symbols. Yang also proposed a new adaptive deep joint source-channel coding wireless image transmission scheme [17]. This scheme introduces a policy network to exploit the trade-off space between rate and signal quality. Zhang proposed a multi-layer semantic-aware communication system for wireless image transmission [18]. The proposed model includes a multi-level semantic feature extractor capable of extracting different levels of semantic information. Kang et al. propose a novel aerial image transfer paradigm for scene classification tasks [19]. A lightweight model for semantic chunk transfer for sensing image and channel conditions is built on a front-end UAV, which achieves a trade-off between transfer latency and classification accuracy by using deep reinforcement learning to explore the semantic patches that best contribute to the back-end classifier under different channel conditions. Dai et al. proposed the NTSCC model [20], which treats nonlinear transformations as vital transformations to efficiently extract source semantic features and provide side information for source-channel coding. Most image transmission systems do not consider the needs of the receiver, but through semantic segmentation, ROI and RONI can be identified and transmitted at different bit rates, thereby achieving the purpose of reducing communication overhead.

3. System model

In this part, a semantic segmentation-based semantic communication system for image transmission is proposed. The system model is shown in Fig. 2, the whole system consists of three parts: the transmitter, the channel and the receiver. The transmitter is used for feature extraction, the channel is used to simulate realistic noise, and the receiver is used for feature recovery to reconstruct the image. We then illustrate the loss function of the system model. The specific details will be described below. Table1 shows the notation table used in this section.

3.1. Transmitter

As can be seen from Fig. 2, the transmitter consists of three separate components: semantic segmentation, semantic encoder and data compression.

Semantic segmentation can classify pixels in an image. In this paper, it is necessary to judge whether the pixel belongs to ROI or RONI. U-Net network is a classic network used for semantic segmentation [21]. The structure of U-Net includes an encoder and a decoder. The encoder extracts features at different levels through multiple downsamsspling, and the decoder fuses multi-level features for decoding. U-Net increases the interaction and fusion of features at different levels, resulting in a stronger feature representation ability, which is a good candidate for our semantic communication system.

After semantic segmentation through U-Net, we can obtain two images, that is, the image of the ROI and

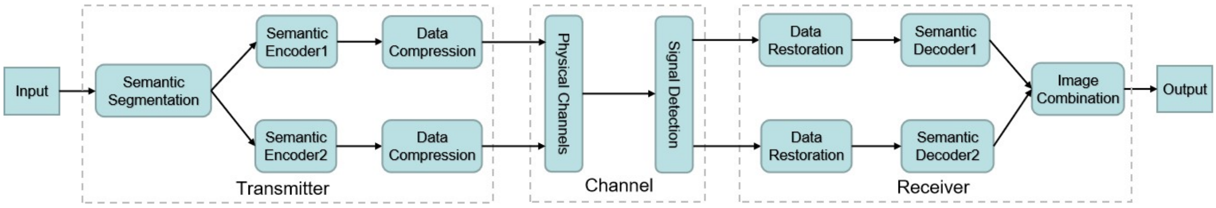


Fig. 2. The proposed system model.

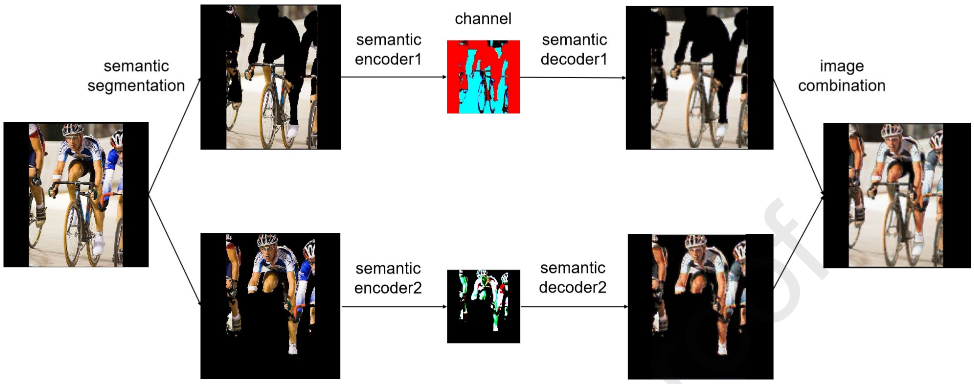


Fig. 3. The result generated by the semantic encoder is sent to the receiver through the channel without data compression.

Table 1
Table of notation.

Variable	Definition
s	the input image
SS_α	semantic segmentation with parameters α
SE_β	semantic encoder with parameters β
SD_ϕ	semantic decoder with parameters ϕ
DC	data compression
D	semantic encoding output
e	element of d
i	index positions of e
N	output of black in semantic encoder
X	result of compression data
$noise$	noise in channel
L	label of each pixel in S
L_{SS}	loss function of semantic segmentation
L_{SC}	loss function of semantic communication

Algorithm 1: Proposed data compression algorithm

Input: Semantic encoding output: D; The most frequent numbers in the semantic encoder output of pure black images: N.

```
foreach (d, n) in (D, N) do
    remove d from n → e;
    index of e in d → i;
    For consecutive values in i, the first and last
      digits are retained;
    e + i → x;
    X + x → X.
```

Output: Result of compression data: X.

the image of the RONI. Since the convolutional neural network is used in the subsequent semantic communication process, the input of the convolutional neural network should be a fixed size. Therefore, we use black to pad in the missing parts of the image to ensure the consistency of the input image size.

The next step is to feed ROI and RONI into semantic communication models with different bandwidths. To this end, we design two semantic communication models, named Semantic Communication 1 (SC1) and Semantic Communication 2 (SC2). Among them, SC1 consists of Semantic Encoder 1 and Semantic Decoder 1, and SC2 consists of Semantic Encoder 2 and Semantic Decoder 2. The model structures of SC1 and SC2 are shown in Table2. As seen from Table2, the model structures of SC1 and SC2 are similar, but SC1 has one more max pooling layer than SC2, so the bandwidth required for SC1 transmission is smaller than SC2. If the result generated by the semantic encoder is sent to the receiver through the channel without data compression, the result is shown in Fig. 3. It means that the results of Semantic Encoder 1 and Semantic Encoder 2 are transmitted simultaneously, leading to increased communication overhead.

Due to the padding of the image, the input image will generate many continuous black areas. Because the convolution computations share the same convolution kernel, the features extracted from the black regions usually have the same color, which means many of the same elements in the feature matrix. If these same elements are treated as 0, the feature matrix can be regarded as a relatively sparse matrix.

Table 2
Parameters of the system model.

	Semantic Communication1		Semantic Communication2	
	Layer Name	Activation / pooling	Layer Name	Activation / pooling
Semantic Encoder	Conv2D	ReLU	Conv2D	ReLU
	Batch Norm	Max pooling	Batch Norm	Max pooling
	Conv2D	ReLU	Conv2D	ReLU
	Batch Norm	Max pooling	Batch Norm	None
Channel	AWGN	None	AWGN	None
Semantic Decoder	ConvTrans2D	ReLU	ConvTrans2D	ReLU
	Batch Norm	None	Batch Norm	None
	ConvTrans2D	ReLU	ConvTrans2D	ReLU
	Batch Norm	None	Batch Norm	None
	ConvTrans2D	ReLU	ConvTrans2D	ReLU

Inspired by the sparse matrix compression algorithm COO [22], we propose a sparse matrix compression algorithm for semantic communication. First, the corresponding elements in the feature matrix can be obtained by inputting a pure black image into the semantic encoder, denoted as n. Then all elements except n are recorded, denoted as e, and the start and end index position i of e are also recorded. The algorithm for matrix compression is given in Algorithm 1. In the transmission process, only e and i need to be passed, without transmitting the entire matrix, which can reduce communication overheads significantly.

Semantic segmentation and semantic encoder are each implemented by a separate neural network. On the transmitter, the image s to be transmitted is input, and then s is divided into s_1 and s_1 by semantic segmentation, where s_1 and s_2 represent the RONI and the ROI, respectively. The data are compressed and then transmitted through the physical channel. Denoting the neural network parameters of semantic segmentation and semantic encoder as α and β, respectively, the encoded symbol sequence x can be expressed as

$$\begin{aligned} x_1 &= DC\left(SE_{\beta 1}\left(SS_{\alpha 1}(s)\right)\right) \\ x_2 &= DC\left(SE_{\beta 2}\left(SS_{\alpha 2}(s)\right)\right) \end{aligned} \tag{1}$$

where $SS_\alpha(\cdot)$ and $SE_\beta(\cdot)$ denote semantic segmentation with parameters α and semantic encoder with parameters β, respectively. Here $DC(\cdot)$ represents the data compression.

3.2. Channel

In the communication channel part, a realistic physical communication environment is simulated. While the real physical communication environment is usually more complicated, for the sake of simplicity, we use the AWGN channel and the rayleigh fading channel that are adopted by many studies [23, 24, 25, 26].

The formula for the SNR is:

$$SNR = 10\lg\frac{P^2}{noise^2} \tag{2}$$

Algorithm 2: Proposed data restoration algorithm

Input: Result of compression data: X; The most frequent numbers in the semantic encoder output of pure black image: n.

foreach x *in* X **do**
  Generate a matrix including all elements of $n \to d$;
  $x \to e + i$;
  e pads in d according to the index of i;
  $D + d \to D$.

Output: Result of restoration data: D.

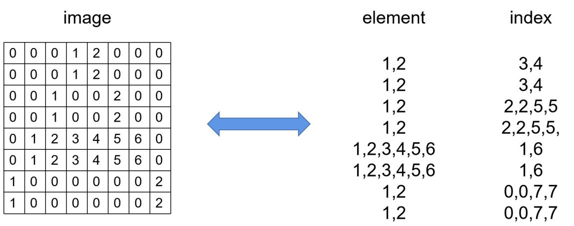


Fig. 4. A simple example of data compression and data restoration.

We set the average power constraint to $P = 1$, and set the SNR of the channel by changing the noise variance $noise$. Therefore, the data received at the receiver are

$$\begin{aligned} x_1' &= x_1 + noise \\ x_2' &= x_2 + noise. \end{aligned} \tag{3}$$

3.3. Receiver

The structure of the receiver is similar to that of the transmitter, and it also consists of three parts, data restoration, semantic decoder and image combination.

Since the input size of the semantic decoder is fixed, the compressed data need to be restored first. The restoration algorithm is shown in Algorithm 2. First, a matrix including all elements of n is generated. The matrix size is the same as the input of the semantic encoder, and n is the corresponding elements in the feature matrix obtained by feeding the solid black image

Algorithm 3: Training process of the proposed system

Input: The dataset S; The label of each pixel in S: L

```
initialization: Initial the weights W and bias b.
while Stop criterion is not met do
    Train the semantic segmentation model.
    Train the semantic communication model.
```

Output: Parameters of the whole network: $\alpha 1, \alpha 2, \beta 1, \beta 2, \phi 1, \phi 2$.

into the semantic encoder. Then the corresponding index position for padding is found according to e and i. In this way, the feature matrix can be restored. Fig. 4 shows a simple example of data compression and data restoration.

After that, semantic decoding is performed according to the feature matrix, and the decoding result is:

$$\begin{aligned} \hat{s}_1 &= SD_{\phi 1}\left(DR\left(x_1^{'}\right)\right) \\ \hat{s}_2 &= SD_{\phi 2}\left(DR\left(x_2^{'}\right)\right) \end{aligned} \tag{4}$$

where $SD_{\phi}(\cdot)$ denotes semantic decoder with parameters ϕ, and $DR(\cdot)$ represents the data restoration.

To reduce the impact of padding part of the data, instead of simply adding two images, the images are combined by finding the maximum value of the pixels:

$$\hat{s} = \max\{\hat{s}_1, \hat{s}_2\} \tag{5}$$

where $\hat{s}$ is the result of the final output.

3.4. Training process

As shown in Algorithm 3, the training process of the proposed system is divided into two stages due to different loss functions. The first stage is to judge whether each pixel belongs to ROI or RONI through semantic segmentation to obtain the input required for the second stage. The second stage is to train the semantic communication network to reduce the information loss of data in semantic communication network. Each stage aims to minimize the loss by gradient descent with mini-batch until the stop criterion is met (the max number of iterations is reached). The two training stages, training of semantic segmentation and training of semantic communication are described in the following sections.

3.4.1. Training of semantic segmentation

The training process of the semantic segmentation model is shown in Algorithm 4. First, minibatch s from the dataset S is used to generate the label l of each pixel, and the loss function of the semantic segmentation network is calculated. The loss function of semantic segmentation can be given by

$$L_{SS} = \sum w(s) \log_{p_{l(s)}} s \tag{6}$$

Algorithm 4: Training process of the semantic segmentation model

Input: The dataset S; The label of each pixel in S: L

```
while Stop criterion is not met do
    BatchSource(S) → s;
    s → l;
    Compute loss L_SS by 6.
    Gradient descent (L_SS).
s, l → s1, s2;
s, s1, s2 → α1, α2;
```

Output: Parameters of the semantic segmentation model: $\alpha 1, \alpha 2$.

Algorithm 5: Training process of the semantic communication model

Input: The ROI s_1; The RONI: s_2

```
Train the semantic communication model of ROI.
while Stop criterion is not met do
    DC(SC_β1(s1)) → x1;
    SD_φ1(DR(x1)) → ŝ1;
    Compute loss L_SC by 7.
    Gradient descent (L_SC).
Train the semantic communication model of RONI.
while Stop criterion is not met do
    DC(SC_β1(s2)) → x2;
    SD_φ1(DR(x2)) → ŝ2;
    Compute loss L_SC by 7.
    Gradient descent (L_SC).
```

Output: Parameters of the semantic communication model: $\beta 1, \beta 2, \phi 1, \phi 2$.

where l is the true label of each pixel and $w(s)$ is a weight map [21]. Then the semantic segmentation model is optimized by the stochastic gradient descent (SGD). This process keeps looping until the stop criterion is met. According to the picture s and the obtained label L, the s_1 and s_s can be calculated.

3.4.2. Training of semantic communication

The training process of the semantic communication model is shown in Algorithm 5. According to Algorithm 4, the s_1 and s_2 after semantic segmentation can be obtained. Since ROI and RONI use different semantic communication models, their parameters must be trained separately. The semantic communication model of RONI is first trained, x_1 is generated by s_1 through semantic encoding and data compression. Then x_1 is data restored and semantic decoded to generate $\hat{s}_1$. The next step is to calculate the loss function. It is desirable to reduce the error between the original input image s and the reconstructed image $\hat{s}$ produced by the decoder. We use the mean squared error (MSE) as the loss function of semantic communication, which is expressed as follows:

$$L_{SC} = \frac{1}{L}\sum_{l=1}^{L}(s - \hat{s})^2. \quad (7)$$

Then the semantic communication model is optimized by the SGD. This process keeps looping until the stop criterion is met. The training process of the semantic communication model of RONI is as same as the ROI.

4. Experimental Evaluation

In this section, we first introduce the experimental setup, which includes datasets, compared schemes, evaluation metrics, and experimental details. Afterward, the results of the experiments are presented and discussed.

4.1. Experimental setup

4.1.1. Datasets

The proposed system has to use a semantic segmentation algorithm which should be trained by labeled pictures in advance. To reduce the workload of labeling image pixels, the semantic segmentation dataset VOC2012 is used to evaluate our system [27], and some preprocessing steps are performed on the VOC2012. The images are classified according to the proportion of different categories in the image, and the category with the largest proportion is the category of the image. Then, images with the category of person are selected as the dataset for this experiment. The region of the person is regarded as ROI, while the rest is regarded as RONI. Since the size of the images in the VOC2012 dataset is not uniform, it is necessary to preprocess the images. Therefore, we resize the image to [512, 512, 3] according to the original scale and pad the insufficient part with black.

4.1.2. Compared schemes

The Compared schemes include using the SC1 model and the SC2 model alone and the model with direct transmission without semantic encoding and decoding, which is defined as traditional communication (TC).

4.1.3. Evaluation metrics

The compression ratio under different algorithms is used as our evaluation index, and the calculation formula is as follows:

$$CR = \frac{\sum_{i=0}^{n} D_i}{\sum_{i=0}^{n} W_i * H_i * C} \quad (8)$$

where D_i is the result of data compression of the ith image, W_i and H_i decibels are the width and height of the ith image, and C is the number of channels of the image, which is 3 here.

Furthermore, a new evaluation metric θPSNR is proposed to measure the performance of the proposed system. The θPSNR is improved based on PSNR(Peak Signal-to-Noise Ratio), the PSNR is an objective standard for evaluating images, and the calculation formula of PSNR is as follows:

$$PSNR = 10 * \lg\frac{MAX^2}{MSE} \quad (9)$$

where MAX is the maximum possible pixel value of the image. The weights θ and (1-θ) of ROI and RONI are considered in θPSNR. The calculation formula of θPSNR is as follows:

$$\theta PSNR = 10 * \lg\frac{MAX^2}{MSE_{roi} * \theta + MSE_{roni} * (1-\theta)} \quad (10)$$

where MSE_{roi} is the mean square error of the ROI, and MSE_{roni} is the mean square error of the RONI.

4.1.4. Model training details

90% of the dataset is used as the training set, and the rest is used as the test set. The Adam optimizer is adopted to train the model [28]. The initial learning rate is 0.01, and the learning rate is halved every 100 epochs for a total of 500. No noise is added during training.

4.2. Experimental results and discussion

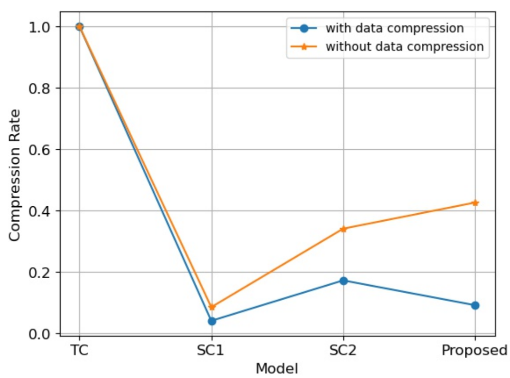
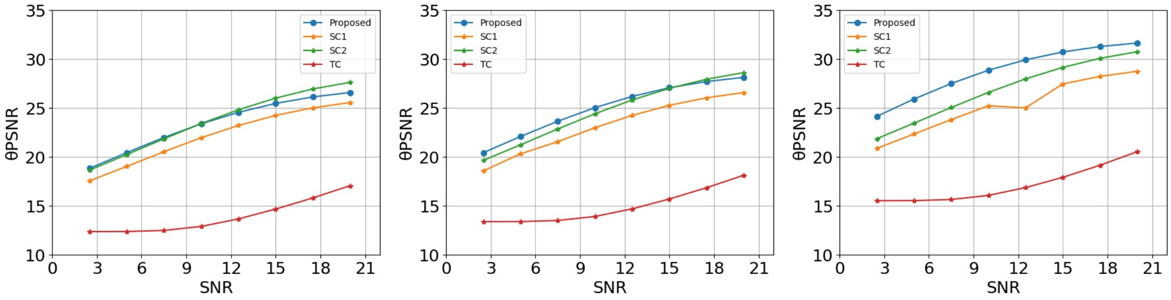


Fig. 5. The compression ratio of different models with and without data compression.

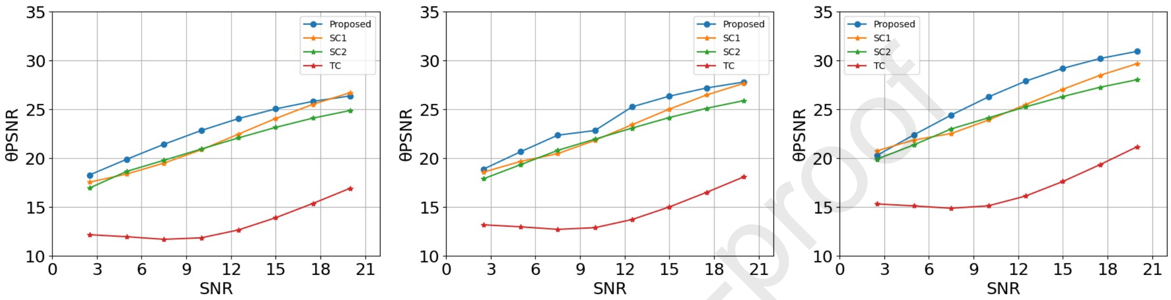
Fig. 5 shows the compression ratio of different models with and without data compression. It can be seen that the compression rate of semantic communication is much better than that of traditional communication, which proves that semantic communication can reduce the overhead required for communication.

Compared with the no data compression model, the compression ratio of SC1 model and SC2 model with data compression is about half. This is because the image input to the semantic encoder needs to be of a uniform size, so the image needs to be completed, which



$\theta = 0.6$ $\theta = 0.8$ $\theta = 1.0$

Fig. 6. the θPSNR under different SNRs over the AWGN channel.



$\theta = 0.6$ $\theta = 0.8$ $\theta = 1.0$

Fig. 7. the θPSNR under different SNRs over the Rayleigh fading channel.

leads to significant communication overhead, and data compression can solve this problem.

Since the proposed system uses both SC1 and SC2, the compression ratio is between SC1 and SC2, about two times that of SC1 and 1/2 that of SC2. However, it is often related to the proportion of ROI in the image for practical applications. The smaller the proportion of ROI in the image, the closer the compression ratio is to SC1, and the larger the ratio is, the closer to the compression ratio of SC2.

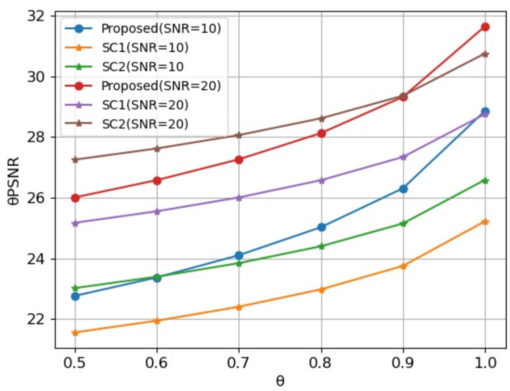


Fig. 8. The θPSNR under different values of θ over the AWGN channel.

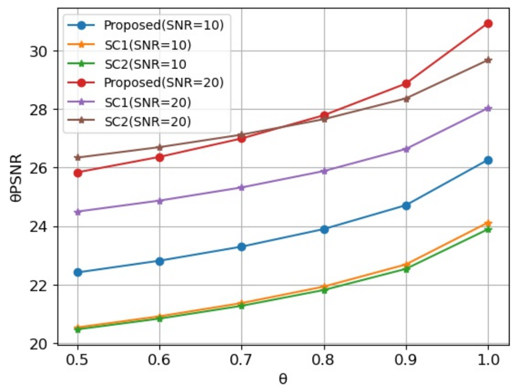


Fig. 9. The θPSNR under different values of θ over the rayleigh fading channel.

Fig. 8 and Fig. 9 show the θPSNR under different values of θ over the AWGN channel and the rayleigh fading channel. When SNR=10, the PSNR of the proposed system is almost better than SC1 and SC2. When SNR=20 and θ=0.5, it indicates the same degree of interest in ROI and RONI. The proposed system's θPSNR is between SC1 and SC2. As θ increases, the θPSNR of the proposed system becomes gradually better than SC2. This means that the higher the receiver's emphasis on ROI, the stronger the superiority of the proposed system.

Next, the θPSNR under different SNRs over the AWGN channel and the Rayleigh fading channel is compared. θ=[0.6, 0.8, 1.0] is chosen and the result is shown in Fig. 6 and Fig. 7. As can be seen, the trans-

Fig. 10. Examples of the transmission results under the SNR = 10.

mission results of semantic communication are better than those of traditional communication, which shows that semantic communication is more robust than traditional communication. The θPSNR value of the proposed system is close to SC2, better than SC1, and far better than TC. With the decrease of SNR, the θPSNR value of the proposed system is changing closer to SC2, and even better than SC2, which shows that the robustness of the proposed system is stronger than that of a single semantic communication network. When θ=1.0, which indicates only ROI is considered, the proposed system outperforms all other models. This is because the images input to the semantic encoder and semantic decoder only have ROI, avoiding the interference of RONI, so the semantic encoder and semantic decoder can better learn the features of ROI. In particular, it can be found that under the Rayleigh channel with low SNR, the performance results of the three semantic communication models are not much different because the heavy influence of Rayleigh fading leads to the loss of semantic information. Therefore, we can know that in the case of low SNR, using low-bandwidth semantic communication to transmit information can achieve a similar performance as high-bandwidth.

Meanwhile, examples of the transmission results under the SNR=10 over the AWGN channel are shown in Fig. 10. The first image is the original input, and the rest of the image are the communication transmission results of TC, SC1, SC2, and the proposed system. As can be seen from Fig. 10, among all the transmission results, the proposed system has the most detailed reconstruction of the person's details and ensures the integrity of the background. For TC, the received images at a low SNR are almost indistinguishable. This is because it does not consider semantics and therefore the transmission results are affected by the noise significantly. By using semantic communication, SC1 and SC2 are less sensitive to noise. The result of SC2 is more detailed than SC1 while requiring a larger bandwidth for transmission. The proposed system achieves the best performance by providing a higher priority for ROI parts in transmissions, which can make an efficient trade-off between performance and cost.

5. Conclusion and future work

This paper proposes a semantic communication system for image transmission based on semantic segmentation. The system can distinguish the ROI and RONI in the image and transmit them in the semantic communication network with different bandwidths so that the information in the ROI can be transmitted more accurately. Furthermore, two semantic communication networks with different bandwidths are designed using convolutional neural networks. Meanwhile, we propose a data compression scheme to reduce communication overhead. Different weights are given to ROI and RONI to evaluate the importance of ROI, and the evaluation metrics of θPSNR are proposed. Compared with single semantic channel and traditional communication networks, our proposed system achieves a notable performance improvement on θPSNR while maintaining a good compression ratio.

Since semantic segmentation and semantic communication are carried out under the process of decoupling, the quality of semantic segmentation would affect the result of semantic communication. How to merge the two networks to reduce the impact of semantic segmentation is one of our future work. On the one hand, since conventional PSNR based approaches evaluate transmission performance based on a simple calculation of differences between the source and received pixels, they are inefficient in representing the quality of semantic communication. We aim to find a better evaluation metric that can measure semantic information in a more effective way.

Acknowledgement

This research was supported in part by ROIS NII Open Collaborative Research 22S0601, and in part by JSPS KAKENHI grant numbers 20H00592 and 21H03424.

References

[1] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, A. P. Sheth, Machine learning for internet of things data analysis: A survey, Digital Communications and Networks 4 (3) (2018) 161–175.

[2] Z. Qin, X. Tao, J. Lu, G. Y. Li, Semantic communications: Principles and challenges, arXiv preprint arXiv:2201.01389 (2021).

[3] Y. Zhou, L. Liu, L. Wang, N. Hui, X. Cui, J. Wu, Y. Peng, Y. Qi, C. Xing, Service-aware 6g: An intelligent and open network based on the convergence of communication, computing and caching, Digital Communications and Networks 6 (3) (2020) 253–260.

[4] C. E. Shannon, A mathematical theory of communication, The Bell system technical journal 27 (3) (1948) 379–423.

[5] H. Xie, Z. Qin, G. Y. Li, B.-H. Juang, Deep learning based semantic communications: An initial investigation, in: GLOBECOM 2020-2020 IEEE Global Communications Conference, 2020, pp. 1–6.

[6] Q. Lan, D. Wen, Z. Zhang, Q. Zeng, X. Chen, P. Popovski, K. Huang, What is semantic communication? a view on conveying meaning in the era of machine intelligence, Journal of Communications and Information Networks 6 (4) (2021) 336–371.

[7] X. Luo, H.-H. Chen, Q. Guo, Semantic communications: Overview, open issues, and future research directions, IEEE Wireless Communications 29 (1) (2022).

[8] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, Understanding convolution for semantic segmentation, in: 2018 IEEE winter conference on applications of computer vision (WACV), Ieee, 2018, pp. 1451–1460.

[9] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, J. Garcia-Rodriguez, A review on deep learning techniques applied to semantic segmentation, arXiv preprint arXiv:1704.06857 (2017).

[10] R. Carnap, Y. Bar-Hillel, An outline of a theory of semantic information, Research Laboratory of Electronics, Massachusetts Institute of Technology (1952).

[11] L. Floridi, Outline of a theory of strongly semantic information, Minds and machines 14 (2) (2004) 197–221.

[12] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, J. A. Hendler, Towards a theory of semantic communication, in: 2011 IEEE Network Science Workshop, 2011, pp. 110–117.

[13] Z. Weng, Z. Qin, G. Y. Li, Semantic communications for speech signals, in: ICC 2021-IEEE International Conference on Communications, 2021, pp. 1–6.

[14] H. Xie, Z. Qin, G. Y. Li, Task-oriented multi-user semantic communications for vqa, IEEE Wireless Communications Letters 11 (3) (2021) 553–557.

[15] H. Xie, Z. Qin, A lite distributed semantic communication system for internet of things, IEEE Journal on Selected Areas in Communications 39 (1) (2020) 142–153.

[16] E. Bourtsoulatze, D. B. Kurka, D. Gunduz, Deep joint source-channel coding for wireless image transmission, IEEE Transactions on Cognitive Communications and Networking 5 (3) (2019) 567–579.

[17] M. Yang, H.-S. Kim, Deep joint source-channel coding for wireless image transmission with adaptive rate control, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 5193–5197.

[18] Z. Zhang, Q. Yang, S. He, M. Sun, J. Chen, Wireless transmission of images with the assistance of multi-level semantic information, in: 18th International Symposium on Wireless Communication Systems, ISWCS, 2022, pp. 1–6.

[19] X. Kang, B. Song, J. Guo, Z. Qin, F. R. Yu, Task-oriented image transmission for scene classification in unmanned aerial systems, IEEE Transactions on Communications 70 (8) (2022) 5181–5192.

[20] J. Dai, S. Wang, K. Tan, Z. Si, X. Qin, K. Niu, P. Zhang, Nonlinear transform source-channel coding for semantic communications, IEEE Journal on Selected Areas in Communications 40 (8) (2022) 2300–2316.

[21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Vol. 9351, 2015, pp. 234–241.

[22] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, H. Van der Vorst, Templates for the solution of linear systems: building blocks for iterative methods, 1994.

[23] Y. Zhang, R. Lu, B. Cao, Q. Zhang, Cooperative jamming-based physical-layer security of cooperative cognitive radio networks: system model and enabling techniques, IET Communications 13 (5) (2019) 539–544.

[24] X. Pang, N. Zhao, J. Tang, C. Wu, D. Niyato, K.-K. Wong, Irs-assisted secure uav transmission via joint trajectory and beamforming design, IEEE Transactions on Communications 70 (2) (2021) 1140–1152.

[25] D. Zhou, K. Niu, C. Dong, Construction of polar codes in rayleigh fading channel, IEEE Communications Letters 23 (3) (2019) 402–405.

[26] H. Xia, K. Alshathri, V. B. Lawrence, Y.-D. Yao, A. Montalvo, M. Rauchwerk, R. Cupo, Cellular signal identification using convolutional neural networks: Awgn and rayleigh fading channels, in: 2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), 2019, pp. 1–5.

[27] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, International journal of computer vision 88 (2) (2010) 303–338.

[28] P. K. Diederik, B. Jimmy, Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations ICLR, 2015.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof