DRL-driven Dynamic Resource Allocation for Task-Oriented Semantic Communication

Haijun Zhang, *Fellow, IEEE,* Hongyu Wang, Yabo Li,
Keping Long, *Senior Member, IEEE*, and Arumugam Nallanathan, *Fellow, IEEE*

Abstract

Semantic communication has been regarded as a promising technology to serve upcoming intelligent applications. However, few studies have addressed the problem of resource allocation in semantic communication networks. Most resource allocation mechanisms act fairly to all original data, ignoring the meaning behind the transmitted bits. In this paper, a dynamic resource allocation scheme for the task-oriented semantic communication network (TOSCN) based on deep reinforcement learning (DRL) is proposed, which allows data with richer semantic information to preferentially occupy limited communication resources. This paper aims to design a deep deterministic policy gradient (DDPG) agent at the micro base station to maximize the long-term transmission efficiency of tasks. Firstly, the relationship between semantic information and task performance is investigated. Subsequently, a novel wireless resource allocation model for TOSCN is proposed by taking the image classification task as an example. Then, a joint optimization problem of the semantic compression ratio, transmit power, and bandwidth of each user is formulated. The agent is trained in an interactive learning environment to obtain a decent trade-off between the amount of data delivered to the receiver and the accuracy of intelligent tasks. Simulation results demonstrate that the proposed scheme achieves significant advantages in relieving communication pressure and improving task performance in resource-constrained wireless networks.

H. Zhang, H. Wang, Y. Li, and K. Long are with the Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, University of Science and Technology Beijing, Beijing 100083, China (e-mails: haijunzhang@ieee.org; hongyuwang@xs.ustb.edu.cn; liyabo@xs.ustb.edu.cn; longkeping@ustb.edu.cn).

A. Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: a.nallanathan@qmul.ac.uk).

Index Terms

Resource allocation, deep reinforcement learning, semantic communication, deep deterministic policy gradient, image classification.

I. Introduction

Under the support of 6G network, the deep integration of Artificial Intelligence (AI) and wireless communication networks has become an inevitable development trend [1]. With the massive connectivity of intelligent devices and the explosion of wireless data traffic, the spectrum scarcity problem has become increasingly prominent, posing huge challenges to wireless communication in the 6G era [2], [3]. The massive data generated by intelligent devices has the characteristic of low value density [4]. However, the current communication technologies focus on the accurate transmission of each symbol, ignoring the target task and the meaning carried in transmission data [5], which results in unnecessary consumption of wireless communication resources. Instead of continuing to pursue the improvement of the network sum-rate, the wireless networks urgently need to make some changes from another perspective to meet the lower latency requirements of emerging intelligent applications.

Recently, the task-oriented semantic communication, which can significantly improve communication efficiency and robustness [6], is expected to become a brand-new communication paradigm in future networks. On the one hand, the task-oriented semantic communication has the ability to extract useful information and remove redundant information for target AI tasks, thereby remarkably reducing the amount of transmitted data and transmission delay. On the other hand, the precise bit recovery is not exacted in semantic communication systems. Thus it is not as susceptible to channel conditions as conventional communication systems. Distinguished from the well-discussed problem of reliable data transmission, introducing the concept of “semantic information” shifts our attention from “how to transmit” to “what to transmit”. Therefore, semantic communication is becoming a superb solution to alleviate the communication bottleneck [7].

There have been some preliminary studies on semantic communication. Aiming at the problem of minimizing the mean square error of image reconstruction tasks, the authors in [8] designed an implicit joint source coding and channel coding scheme. The transmitter and receiver were constructed as symmetric convolutional neural networks (CNNs) at the sending and receiving ends, respectively. Compared to traditional coding methods, the peak signal-to-noise ratio of this image transmission mode showed better robustness when channel conditions became harsh. In [6], an innovative semantic communication system was developed by introducing Transformer in natural language processing yield and successfully used for text transmission. The practical application of Transformer in wireless communication networks faces the dilemma of high model deployment cost and training overhead. Combining the insights from semantic communication with model pruning, the authors in [9] further investigated an affordable semantic communication model for intelligent terminals. The authors in [10] provided a novel compression method for image features, which could decrease the number of feature maps transmitted from smart devices to edge servers and ensure the task success probability of downstream inference. These works lay the groundwork for semantic communication.

Although existing researches have achieved promising results in the design of coding schemes and robust transmissions for semantic communication, very few studies have focused on resource allocation for future semantic communication networks. Most current resource allocation methods take the maximization of energy efficiency or system capacity as the optimization objective and treat the content uploaded by users equally [11], [12]. The ignorance of the specific meaning behind transmitted bits leads to the intense competition of available wireless resources such as bandwidth, power, etc. Considering human perception and user satisfaction, some wireless transmission designs take quality-of-experience (QoE) as an optimization criterion [13]. However, this optimization criterion may not be optimal for machine-to-machine communication scenarios [14], [15]. The traditional communication mode for AI tasks generally transmits all raw data to edge/cloud servers, and thereafter uses pre-trained DL models to acquire inference results. In fact, which part of the data from transmitter is valuable depends on the specific task to be executed [16], [17]. There is often only a small fraction of the data makes a major contribution to the final inference result of the task. Taking pedestrian detection as a simple example, the background and objects other than "pedestrian" in images are not concerned and can be properly compressed due to their almost negligible contribution to the improvement of detection accuracy. If the target task is changed to vehicle detection, only the information with respect to "vehicle" in the image is regarded as valuable, while the information about "pedestrian" becomes redundancy instead. In order to tackle the communication bottleneck and exert the greatest advantages of semantic communication, it is necessary to develop more efficient and appropriate resource allocation schemes that allocate limited communication resources to data with richer semantic information in a task-oriented manner.

Taking semantic information into account for resource allocation, the authors in [18] presented a channel assignment and coding method for text transmission. The authors in [19] designed an adaptive feature compression method to reduce the amount of data to be transmitted. By flexibly controlling compression ratio, a resource allocation mechanism is proposed in [19] to optimize the task success probability. In [20], the authors discussed the performance metric of task-oriented semantic communication network (TOSCN) and defined it as a QoE model. The QoE specifically consists of two components, semantic transmission rate score and semantic similarity score, corresponding to user quality of service and target task performance, respectively. Based on these, a semantic-aware resource allocation method was investigated that maximizes the QoE in TOSCN by optimizing the number of transmitted semantic symbols, channel allocation, and user power. The authors in [21] considered the wireless resource management problem in a heterogeneous network using semantic communication mode, and proposed a new performance metric for this network, named system throughput in message. Then, a heuristic algorithm was applied to solve the problem of user association and bandwidth allocation in the heterogeneous network enabled by semantic communication. The above-mentioned works lay the groundwork for semantic communication and provide useful guidance for the research of this paper.

Most of the existing resource allocation methods for semantic communication focus on the optimization of the short-term network performance. However, there are some scenarios that need to maximize a long-term system gain. In these cases, the loss of short-term gain may promote the whole network to achieve a higher long-term gain. It is challenging to deal with this type of problem using traditional optimization algorithms. Pure data-driven deep reinforcement learning (DRL) has become a powerful tool for solving complex resource management problems in recent years [22]–[24]. By efficiently learning the dynamic changes of the environment, DRL can provide resource allocation strategies that maximize long-term rewards based on pre-trained policy networks. In particular, deep deterministic policy gradient (DDPG) [25] is a kind of model-free and off-policy algorithm with a fast convergence speed. Compared with the value-based Deep Q Network algorithm, DDPG operates over continuous action spaces and directly outputs the optimal allocation strategy without traversing the value function of each action policy, avoiding the problems of excessive quantization error or soaring computational complexity caused by naive discretization [26].

Motivated by the above observations, this paper aims to investigate a resource management mechanism which enables the TOSCN to achieve long-term optimal performance. A system model consisting of multiple semantic communication users and an edge server is considered. Inspired by [19], each user employs the adaptive semantic feature compression approach to control the size of data packets to be delivered to the edge server within a slot. Each user is equipped with a buffer to temporarily store data packets to be transmitted. The transmission efficiency of tasks is defined as the weighted sum of the number of data packets from each user and the corresponding achievable task accuracy at the receiver in a period of time. This paper achieves the maximum transmission efficiency of tasks over a period of time by jointly optimizing the compression ratio and wireless resource allocation strategy of semantic communication users. In this case, a resource allocation strategy that only considers the maximization of the objective function within a single time slot may not be desirable. For example, when the user has more space left in the buffer, a resource allocation scheme that focuses on long-term benefits will avoid giving the user the opportunity to transmit in the current time slot and allocate resources to other users with tight buffers. When encountering the same situation, the scheme that only considers the maximum transmission efficiency of tasks in a single time slot tends to allocate certain resources to each user. This greedy transmission mode increases the degree of compression of semantic features by users, resulting in a decrease in intelligent task accuracy. Therefore, DRL is introduced in this paper to solve the resource allocation problem in TOSCN.

The detailed contributions of this paper are summarized as follows:

- This paper presents a construction method of the background knowledge base (BKB), which stores relationships between semantic compression ratios and AI task performance under various channel states. Take the image classification task as an example, a contribution-based semantic feature compression approach guided by the BKB is investigated.
- A novel wireless resource allocation model for the TOSCN is proposed. A new metric, namely the transmission efficiency of tasks, is defined to measure the network performance from the semantic level. To achieve the preferential occupation of wireless resources by data with richer semantic information, a joint optimization problem of the semantic compression ratio, transmit power, and bandwidth of each intelligent device is formulated.
- With the ultimate goal of maximizing a long-term transmission efficiency of tasks, this paper exploits DRL to tackle the wireless resource management problem in TOSCN. In order to efficiently handle continuous action spaces, a DDPG-driven wireless resource allocation scheme is proposed.

The remainder of this paper is organized as follows. Section II illustrates the resource allocation model. Section III details the proposed DRL-driven dynamic resource allocation scheme for task-oriented semantic communication. The simulation results are presented in Section IV. Finally, Section V provides a brief summary of the research in this paper.

II. System Model and Problem Formulation

A. The Task-Oriented Semantic Communication Model

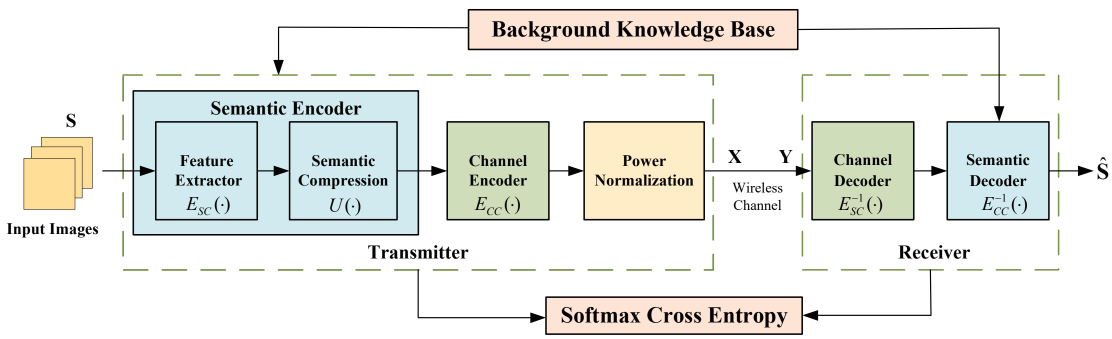


Fig. 1. The architecture of task-oriented semantic communication system.

This paper consider a semantic communication system for image classification task, where the receiver is responsible for feeding back inference results to the transmitter without reconstructing the image. Similar to the traditional communication system framework, the task-oriented semantic communication system includes a transmitter, a wireless channel, and a receiver (shown in Fig. 1). Particularly, traditional source coding is replaced by semantic coding that has the same ability to remove source redundancy. The semantic encoder performs image feature extraction and semantic compression, where the feature extractor consists of the convolutional layers of 18-layer deep residual nets (ResNet18) [27]. An image classifier composed of the fully connected (FC) layer acts as the semantic decoder at the receiver.

The feature extractor performs implicit semantic encoding with the convolutional neural network. For an input image $\mathbf{S}$, the extracted semantic information can be expressed as

$$\mathbf{M}{=}E_{SC}(\mathbf{S},\zeta), \tag{1}$$

where $E_{SC}(\cdot)$ is the semantic extraction network with trainable parameters ζ.

Feature maps are generally regarded as the representations of semantic information in image processing [28]. Each feature map has a different contribution to the correct execution of the task, reflecting the relationship between semantic information and the target AI task. This complex relationship can be naturally represented by model weights in NNs. Assuming that the final inference result of FC layer is z^c, the weight of the i-th feature map with respect to the class label c can be denoted as ω_i^c. Then, ω_i^c can be calculated by using global average pooling and gradient backpropagation as follows

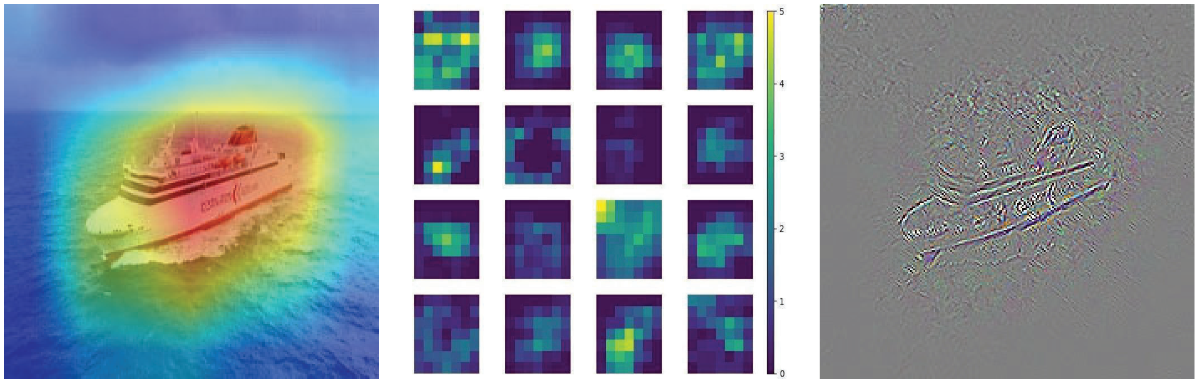
$$\omega_i^c = \overbrace{\frac{1}{W_1 H_1}\sum_m \sum_n \underbrace{\frac{\partial z^c}{\partial F_{m,n}^i}}_{Gradient\ \ Backpropagation}}^{Global\ Average\ Pooling}, \tag{2}$$

where $F_{m,n}^i$ denotes the activation value of the feature map at the *m*-th row and *n*-th column.

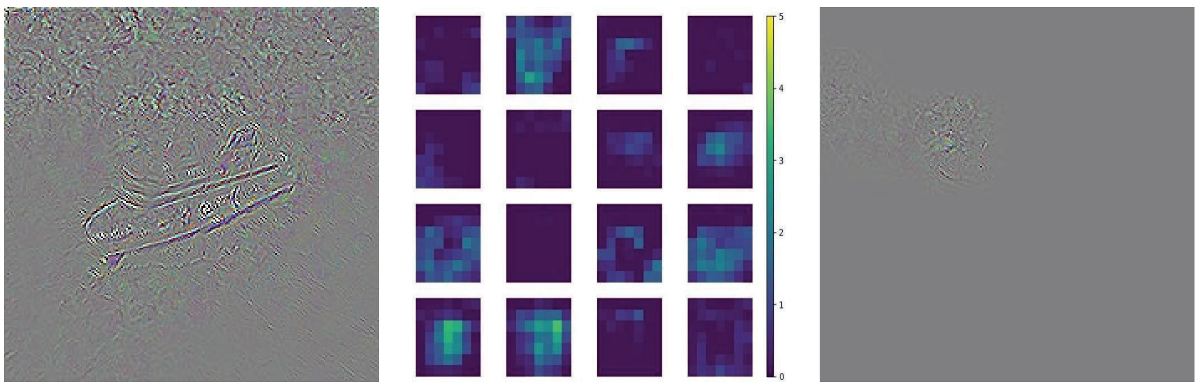
Obviously, ω_i^c greater than zero indicates that the i-th feature map improves the inference probability of class label c. Conversely, the i-th feature map has a reverse effect when ω_i^c is negative. Different from [10], the importance list of feature maps (ILFM) ω^c for class label c can be obtained by taking the absolute value of the weights of feature maps and then sorting them from large to small, which can be denoted as

$$\omega^c = sort(|\omega_1^c|, ..., |\omega_N^c|). \tag{3}$$

The number of feature maps N usually has a large value. However, the few feature maps ranked higher in ILFM actually contain most of the image semantic information, which is sufficient for the identification of the specific object in the image. To prove this viewpoint, a visual explanation derived using the method in [29] is presented in Fig. 2. For the semantic concept "steamship", the gradients flowing into FC layer are combined with extracted feature maps to obtain a coarse-grained class activation map (Fig. 2(a)), displaying the regions that contribute greatly to class discrimination results. In order to further observe the discrepancy of the original image information implied in the feature maps with higher and lower importance scores(Fig. 2(b) and Fig. 2(e)), the gradients at the pixel level are backpropagated. High-resolution visualization results (Fig. 2(c), Fig. 2(d), and Fig. 2(f)) are obtained by dot-multiplying the acquired gradients with the corresponding pixel values. It can be observed that the top 16 feature maps with the highest importance scores contain most of the information in the original image. In resource-restricted and delay-intolerant systems, the feature maps with higher importance score can be given priority to transmission. The feature maps with lower scores can be appropriately discarded to achieve the purpose of cutting back the wireless communication cost. In this paper, the above



(a) The class activation map. (b) Top 16 feature maps with the highest importance score. (c) Image information contained in (b).



(d) Image information contained in all feature maps. (e) The 16 feature maps with the lowest importance score. (f) Image information contained in (e).

Fig. 2. The image information implicit in feature maps.

operations are defined as semantic compression. After semantic compression, the data fed into the channel encoder can be denoted as

$$\mathbf{O} = U(\mathbf{M}, \eta), \tag{4}$$

where $U(\cdot)$ represents the semantic compression operation, whose calculation process can be denoted as

$$U(F^i, \eta) = \begin{cases} F^i, & |\omega_i^c| \geq \omega_\eta \\ \mathbf{0}, & |\omega_i^c| < \omega_\eta \end{cases}, \tag{5}$$

where F^i is the i-th feature map, ω_η is the compression threshold, and $\eta \in [0, 1)$ denotes the semantic compression ratio.

Next, the channel encoder maps the compressed data into symbols suitable for transmission over the wireless channel, which can be denoted by

$$\mathbf{X}' = E_{CC}(\mathbf{O}, \theta), \tag{6}$$

where $E_{CC}(\theta)$ is the channel encoder network with trainable parameters θ.

To meet transmit power constraints, the data actually sent to the physical channel should be further normalized as

$$\mathbf{X} = \frac{\mathbf{X}' \times \sqrt{dim(\mathbf{X}') \times P}}{\|\mathbf{X}'\|_2}, \tag{7}$$

where $dim(\mathbf{X}')$ denotes the dimension of the vector $\mathbf{X}'$.

Inevitably, semantic compression leads to a decline in task performance, therefore, how to find the right compression ratio to achieve an optimal trade-off between transmission costs and semantic correctness is the most critical issue in the wireless resource allocation of semantic communication. Based on the ILFM corresponding to different AI tasks, the mathematical relationship between compression ratio and AI task performance is studied and stored in the BKB shared by the semantic encoder and decoder. The subsequent resource allocation is instructed by the constructed BKB, whose detailed process will be discussed in *subsection B*.

After passing through the physical channel, the data received by the receiver can be represented as $\mathbf{Y}$. Then the output $\hat{\mathrm{M}}$ after channel decoding at the edge server is given by

$$\hat{\mathbf{M}}{=}E_{SC}^{-1}(\mathbf{Y}, \chi), \tag{8}$$

where $E_{SC}^{-1}(\cdot)$ is the channel decoder network with trainable parameters χ.

The semantic decoder is responsible for converting the data output by the channel decoder into a series of probability values, and infers the result of image classification according to the maximum probability value. Therefore, the final semantic restoration result corresponding to the original input image can be obtained by

$$\hat{\mathbf{S}} = E_{CC}^{-1}(\hat{\mathbf{M}}, \delta), \tag{9}$$

where $E_{CC}^{-1}(\cdot)$ is the semantic decoder network with trainable parameters δ.

To minimize semantic errors, the softmax cross-entropy (CE) is used to characterize the difference in probability distributions between the ground-truth labels of input images and outputs. Considering the image classification problem with a label set $[C] = \{1, 2, ..., C\}$ and an instance $\mathbf{S}$, the output of the last FC layer can be denoted as $l_p = [l_p^1, ..., l_p^C]$. Then the one-hot encoding corresponding to the ground-truth label of $\mathbf{S}$ can be denoted as $l_g = [l_g^1, ..., l_g^C]$. With regard to class-balanced samples, the loss function for training the transmitter-receiver can be expressed as

$$L_{CE} = -\sum_{c=1}^{C} l_g^c \cdot \log\left(\frac{e^{l_p^c}}{\sum_{d=1}^{C} e^{l_p^d}}\right) = \sum_{c=1}^{C} l_g^c \cdot \log\left(1 + \sum_{d=1, d\neq c}^{C} e^{l_p^d - l_p^c}\right). \tag{10}$$

When encountering the problem that the image category labels exhibit an imbalanced or long-tailed distribution [30], the loss function can be adjusted by introducing class prior probability as follows

$$L_{CE} = -\sum_{c=1}^{C} l_g^c \cdot \log\left(\frac{e^{l_p^c + \rho\cdot\log p(c)}}{\sum_{d=1}^{C} e^{l_p^d + \rho\cdot\log p(d)}}\right) = \sum_{c=1}^{C} l_g^c \cdot \log\left(1 + \sum_{d=1, d\neq c}^{C} \left(\frac{p(d)}{p(c)}\right)^{\rho} \cdot e^{l_p^d - l_p^c}\right), \tag{11}$$

where $\rho\cdot\log p(d)$ and $\rho\cdot\log p(c)$ denote the label-dependent offsets of label d and c, respectively. ρ is a positive constant with a suitable value. $p(d)$ and $p(c)$ denote the empirical class frequencies of label d and c, respectively.

B. Resource Allocation Model

In this part, a novel wireless resource allocation model for TOSCN is considered. A joint optimization problem of the semantic feature compression ratio, transmit power, and bandwidth of each intelligent device is formulated. The proposed resource allocation scheme can be easily expanded to different AI tasks, and the image classification task is mainly discussed in this paper. In the NN model, the number of parameters of FC layer accounts for the majority. Therefore, a distributed semantic communication network that deploys FC layer to the edge server is considered to make devices affordable. Specifically, the communication process includes the following four steps:

1) Intelligent devices sequentially perform feature extraction, semantic compression and channel encoding for captured images based on BKB, and then generate a corresponding data packet for each image.

2) The data packets are uploaded to the edge server.

3) The edge server perform intelligent processing and computation according to the trained model.

4) The inference results of AI tasks are fed back to the corresponding devices for subsequent processing.

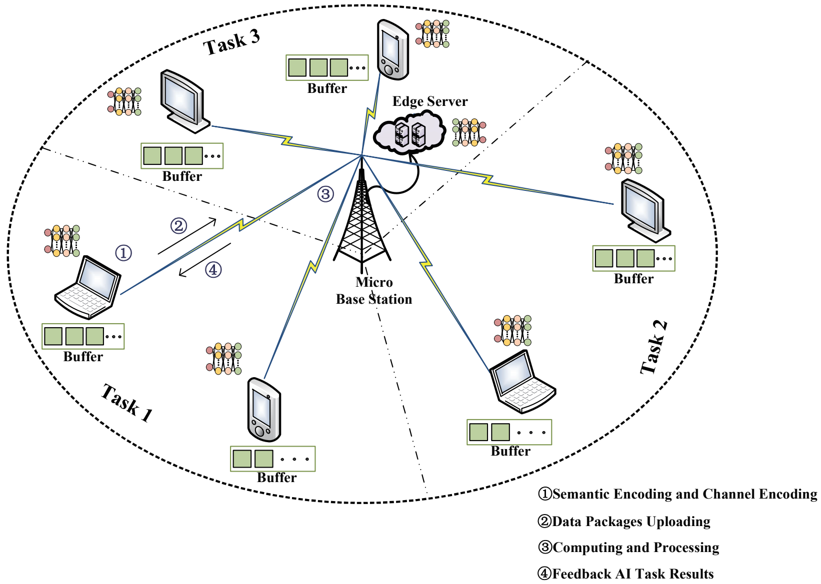


Fig. 3. The task-oriented semantic communication scenario in this paper.

As illustrated in Fig. 3, there are D intelligent devices and an edge server in the system model. Consider a resource scheduling period with T time slots, every slot has a duration of L. Each device performs a specified image classification task over a period of time, such as surface defect classification, commodity classification, etc. Supposed that the number of task categories is J and the number of devices to perform task j is n_j, it is easy to obtain $\sum_{j=1}^{J} n_j = D$. Each user is equipped with a vision sensor and performs semantic feature extraction on the captured images accordinn g to their respective processing speed. The extracted semantic features will be temporarily stored in a buffer with a maximum capacity v_{max}. If the buffer is full, the device will stop processing images until the storage is released.

After the feature extraction, semantic compression and channel encoding of an image, the data stream actually transmitted on the channel is defined as a data packet. Without considering semantic compression, the data packets generated by feature extraction and channel encoding have the same size b for all users. It is a reasonable assumption since images are typically resized to a fixed height and width before feature extraction. Denoting the i-th user corresponding to task j as $u^{i,j}$, the size of a single data packet sent by user $u^{i,j}$ in the t-th time slot can be written as

$$\hat{b}_t^{i,j} = (1 - \eta_t^{i,j})b, \tag{12}$$

where $\eta_t^{i,j}$ denotes the compression ratio of user $u^{i,j}$.

During the t-th slot, the transmission rate of user $u^{i,j}$ can be calculated by

$$R_t^{i,j} = B_t^{i,j} \log(1 + \frac{P_t^{i,j} h_t^{i,j}}{\sigma_t^{i,j^2}}), \tag{13}$$

where $B_t^{i,j}$ and $P_t^{i,j}$ denote the bandwidth and transmit power assigned to user $u^{i,j}$, respectively. σ_t^{i,j^2} and $h_t^{i,j}$ are the power of additive white Gaussian noise and channel gain between user $u^{i,j}$ and the receiver, respectively. Denoting the noise power per unit bandwidth as N_0, the received noise power from user $u^{i,j}$ can be expressed as

$$\sigma_t^{i,j^2} = N_0 B_t^{i,j}. \tag{14}$$

The channel gains are represented as independent random variables while considering both large-scale fading as well as small-scale Rayleigh fading. It is supposed that the gain of each channel remains constant within a single slot interval and varies independently from slot to slot. In the t-th slot, the channel gain $h_t^{i,j}$ between user $u^{i,j}$ and the receiver can be described as

$$h_t^{i,j} = \alpha^{i,j} g_t^{i,j}, \tag{15}$$

where the large-scale fading part $\alpha^{i,j}$ can be further expressed as

$$\alpha^{i,j} = G^{i,j} \beta^{i,j} (d^{i,j})^{-\varphi^{i,j}}, \tag{16}$$

where $G^{i,j}$ denotes the pathloss constant, $\beta^{i,j}$ is the shadowing component which obeys logarithmic normal distribution, $d^{i,j}$ is the distance from user $u^{i,j}$ to the receiver, and $\varphi^{i,j}$ is the pathloss exponent.

The small-scale fading part $g_t^{i,j}$ is time-varying and can be modeled as a first-order complex Gauss-Markov process as follows

$$g_t^{i,j} = \rho(L) g_{t-1}^{i,j} + e_t^{i,j} \sqrt{1 - \rho^2(L)}, \tag{17}$$

where $\rho(L) = J_0(2\pi f_d L)$ denotes the autocorrelation function which is dependent on the maximum Doppler frequency f_d and used to measure the correlation between two successive fading blocks. $J_0(.)$ denotes the zeroth-order Bessel function. $e_t^{i,j}$ denotes a circularly symmetric complex Gaussian random variable with the unit variance.

It is assumed that the intelligent devices have parallel computing and processing capabilities to encode the packets queued in buffers and waiting to be sent in advance. For simplicity, the data collection, data encoding, and data transmission can be roughly considered as three independent processes [31]. Therefore, the number of data packets that user $u^{i,j}$ can transmit in slot t satisfies the following equation

$$v_t^{i,j} = \frac{LR_t^{i,j}}{\hat{b}_t^{i,j}}. \tag{18}$$

Considering the actual transmission scenario and the maximum buffer capacity, the actual number of packets delivered to the base station can be denoted as

$$v_t^{i,j} = \min\left\{ \left\lfloor \frac{LR_t^{i,j}}{\hat{b}_t^{i,j}} \right\rfloor, \hat{v}_t^{i,j} \right\}, \tag{19}$$

where $\lfloor . \rfloor$ denotes the flooring operation, and $\hat{v}_t^{i,j}$ is the existing quantity of packets in the buffer belongs to user $u^{i,j}$ at the start of the t-th slot.

Assuming that the number of packets accumulated by user $u^{i,j}$ during the t-th slot is $\bar{v}_t^{i,j}$, the existing quantity of packets at the start of the $(t+1)$-th slot can be denoted as

$$\hat{v}_{t+1}^{i,j} = \min\{\hat{v}_t^{i,j} - v_t^{i,j} + \bar{v}_t^{i,j}, v_{\max}\}. \tag{20}$$

Different from traditional communication, TOSCN can prioritize data with higher contribution in resource allocation. In the previous subsection, we introduced a method to quantify the contribution of different semantic features. The question that arises naturally is how to make the sender and receiver acquire prior knowledge about the impact of contribution-based semantic compression on AI task performance. Therefore, a BKB shared by the sender and receiver needs to be constructed to instruct resource allocation. In fact, the accuracy of image classification at the receiver is affected by both the semantic compression ratio and channel state. Regarding the variation of the image classification accuracy with the channel state, there is currently no specific mathematical expression. Fortunately, modeling the physical channel as a non-trainable fully connected layer can simulate different channel states. Drawing support from the curve fitting method, the mathematical relationships between compression ratios and task performance in various channel states are explored. Based on the previously collected ILFM, the effect of the semantic compression ratio on the classification accuracy of different tasks is evaluated. For user $u^{i,j}$, the mathematical characterization between classification accuracy $A_t^{i,j}$ and semantic compression ratio $\eta_t^{i,j}$ under a fixed channel state can be modeled as follows

$$A_t^{i,j} = \alpha_1^j (\eta_t^{i,j})^{\alpha_2^j} + \alpha_3^j, \tag{21}$$

where the value range of i is $[1, n_j]$. α_1^j, α_2^j, and α_3^j are the parameters corresponding to task j. The mean square error of the prediction accuracy and the actual accuracy is used as the loss function, and the Levenberg-Marquardt method is employed to minimize the loss function and solve the three parameters.

In TOSCN, the goal of resource management should be closely related to intentions. For intelligent tasks that take into account the quality of human experience, the raw data needs to be reconstructed at the receiver for human viewing. This means that the probability of a task being performed correctly is not the only factor needs to be optimized. At this point, more communication resources must be used to improve the visibility of reconstructed data, such as image clarity. For machine-to-machine communication, it is only necessary to consider whether the automated task can be performed correctly and how efficiently it is performed. By comprehensively considering the classification accuracy as well as the number of packets successfully sent in a period, this paper uses the transmission efficiency of tasks as a metric to verify the performance of the proposed resource allocation scheme. The transmission efficiency of tasks is defined as the weighted sum of the number of data packets from each user and the corresponding achievable task accuracy at the receiver. Specifically, the transmission efficiency of tasks v_t in slot t is defined as follows

$$v_t = \sum_{j=1}^{J} \sum_{i=1}^{n_j} v_t^{i,j} \times A_t^{i,j}. \tag{22}$$

Obviously, the increase of the semantic compression ratio is capable of reducing data to be transmitted, thereby decreasing the occupied bandwidth of users and the required transmission delay. Nonetheless, the lossy compression of semantic features inevitably brings about a drop in classification accuracy. For a decent trade-off between the quantity of data packets delivered to the receiver and the accuracy of intelligent tasks, a maximization problem is formulated to simultaneously optimize the compression ratio, transmit power, and bandwidth of each user equipment (UE) according to the available wireless resources. The ultimate goal of this paper is maximizing a long-term transmission efficiency of tasks. Based on the above assumptions, the corresponding optimization problem can be written as

$$\max_{B_t^{i,j},\, P_t^{i,j},\, \eta_t^{i,j}} \sum_{t=1}^{T} v_t \tag{23}$$

$$s.t. \sum_{j=1}^{J}\sum_{i=1}^{n_j} B_t^{i,j} \leq B_{\max}, \forall t, \tag{23a}$$

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j} P_t^{i,j} \leq P_{\max}, \forall t, \tag{23b}$$

$$B_t^{i,j} \geq 0, \forall i, j, t, \tag{23c}$$

$$P_t^{i,j} \geq 0, \forall i, j, t, \tag{23d}$$

$$A_t^{i,j} \geq A_{\min}, \forall i, j, t, \tag{23e}$$

where constraints (23a)-(23d) ensure that the allocated resources for bandwidth and transmitted power are non-negative and no more than their limits. Constraint (23e) restricts the predicted classification accuracy of each user to be no lower than $A_{\min}$.

In the above problem, the loss of short-term gain may promote the whole network to achieve higher long-term gains. Accordingly, a model-free DRL algorithm is used, which will be discussed in next section.

III. DRL-driven Resource Allocation Scheme

Due to the powerful decision-making capability, DRL has been widely applied in resource allocation such as user association and power control in recent years. In this section, a DRL-based dynamic resource allocation scheme for TOSCN is developed. The DDPG framework is employed to acquire the sensible solution of the considered optimization problem.

A. The DDPG Framework

A standard DRL setup involving an agent that observes the noisy environment in discrete time slots is considered. The DDPG agent interacts with the dynamic environment to obtain the state, then it is input to the action network to get the bandwidth and power allocation strategy as well as compression scheme for data sent by each user. After executing the action policy, the agent will acquire feedback from environment and assess the value of the policy to optimize the parameters of NNs.

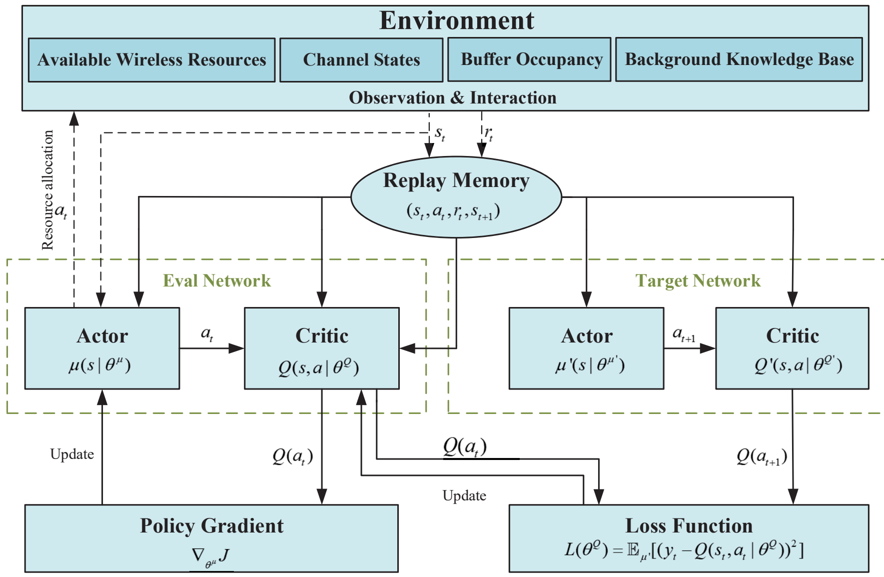


Fig. 4. The DDPG framework.

The framework of DDPG is given in Fig. 4. DDPG consists of an actor network and a critic network as well as their respective copies, namely, target actor network and target critic network, which get parameters from the actor or critic to soft-update its own parameters. This paper use $\mu(s|\vartheta^{\mu})$ with parameter ϑ^{μ}, $\mu'(s|\vartheta^{\mu'})$ with parameter $\vartheta^{\mu'}$, $C(s,a|\vartheta^{C})$ with parameter ϑ^{C}, and $C'(s,a|\vartheta^{C'})$ with parameter $\vartheta^{C'}$ to denote the actor network, target actor network, critic network, and target critic network, respectively. Specifically, the action network $\mu(s|\vartheta^{\mu})$ selects action a_t based on current state s_t and behavior noise $\mathcal{N}_t$ in each interaction, which can be expressed as

$$a_t = \mu(s_t|\vartheta^{\mu}) + \mathcal{N}_t, \tag{24}$$

where $\mathcal{N}_t$ obeys the Gaussian distribution with mean μ_e and variance σ_e^2. Note that the behavior noise will only be added to the actions determined by the actor network in training stage, and it is not needed in the inference stage and the update stage of network parameters. The introduction of behavior noise increases the likelihood of finding better policies.

At time slot t, the agent will acquire an instant reward r_t after performing a_t and thereafter observing the next state s_{t+1}. The critic network is an approximator of action-value function, which describes the expectation of total discounted future reward. Assuming that the discount factor is denoted by β, the return reward G_t from slot t to the end of the iteration T can be denoted as follows

$$G_t = r_t + \beta r_{t+1} + \beta^2 r_{t+2} + ... + \beta^{T-i} r_T = \sum_{i=t}^{T} \beta^{(t-i)} r_t. \tag{25}$$

Therefore, the action-value function after the execution of a_t in state s_t by following the deterministic policy $\mu(s|\vartheta^{\mu})$ can be written as

$$C(s_t, a_t) = \mathbb{E}_{r_{i\geq t}, s_{i>t}, a_{i>t}}(G_t|s_t, a_t). \tag{26}$$

As in DQN, DDPG stores previous transitions as tuples (s_t, a_t, r_t, s_{t+1}) in a fixed-capacity experience replay buffer denoted by R. In the process of training the policy networks, tuples from the replay buffer are randomly sampled to break potential associations between transitions produced by exploration with time continuation. Denoting the target action-value of a sample is y_t, it is given by

$$y_t = r_t + \beta C'(s_{t+1}, \mu'(s_{t+1}|\vartheta^{\mu'})|\vartheta^{C'}). \tag{27}$$

Denoting the sampling batch size as N, the parameters of critic network ϑ^C are updated using gradient backpropagation with a loss function of

$$L(\vartheta^C) = \frac{1}{N} \sum_{t=1}^{N} (y_t - C(s_t, a_t|\vartheta^C))^2. \tag{28}$$

The objective function $\mathcal{J}_{\psi}$ is a metric of the effect of the action network $\mu(s|\vartheta^{\mu})$, which can be defined as

$$\mathcal{J}_{\psi} = \mathbb{E}_{s\sim\psi}[C(s, \mu(s|\vartheta^{\mu}))], \tag{29}$$

where ψ denotes the state distribution function. The final objective of training the DDPG framework is to seek an optimal action network to maximize $\mathcal{J}_{\psi}$, which can be expressed as

$$\mu(s|\vartheta^{\mu}) = \arg\max \mathcal{J}_{\psi}. \tag{30}$$

The update of action network can be achieved by applying the chain rule to the sampled performance objective function as follows

$$\nabla_{\vartheta^{\mu}} \mathcal{J}_{\psi} = \frac{1}{N} \sum_{t=1}^{N} \nabla_a C(s, a|\vartheta^C)|_{s=s_t, a=\mu(s_t|\vartheta^{\mu})} \nabla_{\vartheta^{\mu}} \mu(s|\vartheta^{\mu})|_{s=s_t}. \tag{31}$$

Consequently, the process of updating ϑ^{μ} by gradient descent can be expressed as

$$\vartheta^{\mu} = \vartheta^{\mu} - \alpha_{actor} \nabla_{\vartheta^{\mu}} \mathcal{J}_{\psi}, \tag{32}$$

where α_{actor} denotes the learning rate of action network.

Directly copying the weight parameters of eval networks to the two target networks will lead to large fluctuations in the loss function. In order to maintain the stability of learning, the parameters of target networks, namely, $\vartheta^{\mu'}$ and $\vartheta^{C'}$, are soft-updated according to the update coefficient ε as follows

$$\vartheta^{\mu'} = \varepsilon\vartheta^{\mu} + (1-\varepsilon)\vartheta^{\mu'}, \tag{33}$$

$$\vartheta^{C'} = \varepsilon\vartheta^{C} + (1-\varepsilon)\vartheta^{C'}. \tag{34}$$

B. The Concrete DRL Design

To train the DDPG agent, simulated physical environments with time-varying channel states, background knowledge corresponding to target tasks, buffer occupancy and the total available wireless resources at the base station are constructed. The concrete state, action, and reward function are defined as below:

1) *State Space:* In TOSCN, the state space is jointly determined by communication environment, task assignment, and the buffer occupancy of users. The system state at slot t can be denoted as the following tuple

$$s_t = \{n_1, ..., n_J, h_t^1, ..., h_t^D, \hat{v}_t^1, ..., \hat{v}_t^D\}, \tag{35}$$

where J is the number of classification task categories, and n_j is the number of devices to perform task j. $n_1, ..., n_J$ are discrete variables that depend on the task assignment. $h_t^1, ..., h_t^D$ denote the channel gains from users to the base station at slot t. $\hat{v}_t^1, ..., \hat{v}_t^D$ denote the queue length of users at the beginning of current slot.

2) *Action Space:* The agent directly maps the current state s_t to an action a_t which includes the compression ratio, bandwidth proportion, and power proportion of each user. Specifically, the action a_t can be defined as

$$a_t = \{\eta_t^1, ..., \eta_t^D, B_t^1, ..., B_t^D, P_t^1, ..., P_t^D\}. \tag{36}$$

The output of actor network in DDPG is a set of continuous values. Supposing that the total number of features of each image is F, for user u in the scenario, the number of discarded feature maps can be calculated by $\lceil \eta_t^u F \rceil$, where $\lceil . \rceil$ denotes the ceiling operation. Similarly, the actually allocated bandwidth and power for user u can be obtained by $\lfloor B_t^u B_{\max} \rfloor$ and $\lfloor P_t^u P_{\max} \rfloor$, where $\lfloor . \rfloor$ denotes the flooring operation. The softmax function is applied to the output action $B_t^1, ..., B_t^D$ and $P_t^1, ..., P_t^D$ to satisfy the constraints (23a)-(23d).

3) *Reward:* The agent aims to achieve the maximum improvement in the transmission efficiency of tasks, therefore, a higher value of objective function is expected. Without violation of the constraint (23e), the instant reward is naturally defined as the the transmission efficiency of tasks in current slot t. To further cut down the training overhead of the DDPG agent and improve the performance of the proposed scheme, the agent will be punished when constraint (23e) is not satisfied. The reward function can be expressed as

$$r(s_t, a_t) = \begin{cases} v_t, & if \ A_t^u \geq A_{\min}, \forall u, \\ A_t^u - A_{\min}, & if \ A_t^u < A_{\min}, \exists u. \end{cases} \tag{37}$$

4) *State Normalization:* A long-standing problem in reinforcement learning is that the distribution of the input data affects the output of the activation function [32]. As far as the tanh function is concerned, an excessively large or excessively small input value will locate the result in the saturated part of the activation function, causing the output value is infinitely close to 1 or -1. This phenomenon makes it difficult to update the parameters of NNs by gradient descent method. The observed states are preprocessed by batch normalization to narrow the variation range of inputs, which can more effectively utilize the sensitive part of the activation function to non-linearize the data with different physical units. To handle the magnitude difference of variables in the state set, two scaling factors φ_D and φ_v are introduced in the proposed algorithm to scale down $n_1, ..., n_J$ and $\hat{v}_t^1, ..., \hat{v}_t^D$, which are respectively equal to the maximum values of the corresponding variables, namely, φ_t=$v_{\max}$ and φ_D=D.

The detailed process of solving problem (23) is given in Algorithm 1.

Algorithm 1 DDPG-driven Agent Training for Resource Allocation

Input:

Episode length E, step length T, discount factor β, the soft-update coefficient ε, actor learning rate α_{actor}, critic learning rate α_{critic}, the capacity of experience buffer N_M, batch size N, the mean value μ_e, standard deviation σ_e of Gaussian distributed exploration noise $\mathcal{N}$, the number of tasks categories J, total available power $P_{\max}$, total available bandwidth $B_{\max}$, minimum classification accuracy $A_{\min}$, maximum queue length for packets $\upsilon_{\max}$, and total number of UEs D.

Output:

The actor network $\mu(s|\vartheta^{\mu})$ that determines resource allocation strategy.

Initialize the parameters ϑ^{μ}, $\vartheta^{\mu'}$, ϑ^{C}, and $\vartheta^{C'}$ of the four neural networks.
for $e = 1, 2, ..., E$ **do**
    Reset the task assignment and user distribution.
    Normalize the initial state to obtain s_1.
    for $t = 1, 2, ..., T$ **do**
        Get a resource allocation policy with current actor network and the behavior noise $\mathcal{N}$.
        Perform action a_t, compute instant reward $r(s_t, a_t)$, and observe next state.
        Normalize next state and get s_{t+1}.
        if the experience replay buffer does not overflow **then**
            Cache tuple (s_t, a_t, r_t, s_{t+1}) in the experience buffer R.
        else
            Randomly replace a tuple in R.
            Randomly sample N tuples from R.
            Get the target action-value via (27).
            Update ϑ^{C} via minimizing the loss function in (28).
            Update ϑ^{μ} via (31) and (32).
            Soft-update $\vartheta^{\mu'}$ and $\vartheta^{C'}$ via (33) and (34).
        end if
    end for
end for

IV. Numerical Results

In this section, numerical simulations are shown to verify the advantage of the proposed DDPG-based dynamic resource allocation scheme for task-oriented semantic communication. A microcell with 6 UEs and a micro base station is considered. Each UE is instructed to recognize objects in the captured images for subsequent processing. The detailed parameters of the DDPG-based DRL framework are listed in Table I, unless specified.

TABLE I
SIMULATION PARAMETERS.

Parameter Name	Value	Parameter Name	Value
The number of intelligent devices, D	6	The learning rate of actor, α_{actor}	0.0003
The number of task types, J	3	The learning rate of critic, α_{critic}	0.0005
Maximum capacity of buffers, $v_{\max}$	6	Soft update coefficient, ε	0.01
Minimum classification accuracy, $A_{\min}$	0.6-0.8	The capacity of the replay buffer, N_M	150
Initial data size, b	0.2 M	Sample batch size, N	64
Total system transmit power, $P_{\max}$	0.14 -0.2 W	Iteration times, E	600
Cell radius, r	100 m	The step size of DDPG, K	10
Effective thermal noise power, N_0	-174 dBm/Hz	The mean of behavior noise, μ_e	0
Total system bandwidth, $B_{\max}$	2 MHz-8 MHz	The standard deviation of behavior noise, σ_e	0.1
Length of time slot, L	200 ms	Discount factor, β	0.9
Path loss	128.1+37.6$\log_{10}(d)$	The standard deviation of shadow fading	6 dB

TABLE II
THE NEURAL NETWORK ARCHITECTURE FOR IMAGE CLASSIFICATION TASKS.

Parameter	Layer Name	Output Size	Activation Function
Transmitter	Conv2d	112×112, 64	Relu
	ResNet18 Block 1	56×56, 64	Relu
	ResNet18 Block 2	28×28, 128	Relu
	ResNet18 Block 3	14×14, 256	Relu
	ResNet18 Block 4	7×7, 512	Relu
Channel	FC Layer	None	None
Receiver	FC Layer	10	Softmax

The neural network architecture for image classification tasks is given in Table II. The adopted image dataset are MNIST, Fashion-MNIST, and CIFAR-10, corresponding to task 1, task 2, and task 3, respectively. For the sake of mitigating the impact of the randomness introduced by the physical channel, all images are transmitted 10 times.

Firstly, the robustness of the proposed semantic communication method and baseline transmission methods (JPEG coding) to variations of the average channel signal-to-noise ratio (SNR) is investigated in Fig. 5. The semantic communication model is trained with an average SNR of

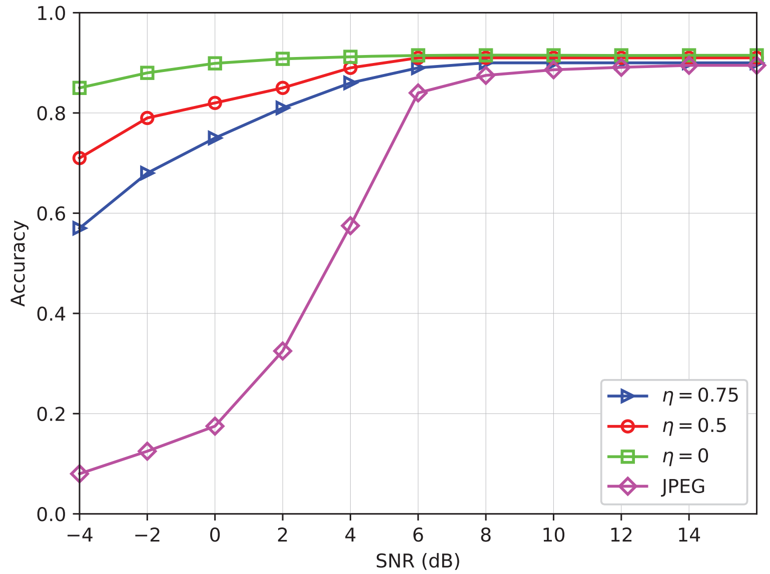


Fig. 5. Classification accuracy versus SNR, with the feature map transmission scheme using semantic encoding and compression and the raw image transmission scheme using JPEG encoding.

13 dB and a learning rate of 10^{-4}. Consistent with the anticipation, the classification accuracy of all these methods shows a significant downward tendency with low SNR (about less than 5 dB). Specifically, a higher degree of compression leads to a larger performance penalty in low SNR regime. The overall performance of the DL-driven semantic communication method is far superior to that of traditional communication method which suffer from the "cliff effect" due to direct transmission of raw images without semantic-level processing. Most importantly, when the channel quality drops below 2 dB, the encoded image using JPEG can hardly complete the task, while the proposed method still maintains a good robustness and displays a graceful degradation of the accuracy. The reason for the stronger noise immunity of the latter is that the values in the feature maps extracted by CNN have a sparse distribution. Similar to the conclusions in [8], the communication mode that transmits feature maps can achieve better task performance when the actual SNR is around the training SNR. Despite being trained at a fixed SNR, the encoded representations of images learned by our model exhibit a good resilience to the fluctuations in channel quality. This characteristic is of great significance for data transmission in time-varying channels or communication using multiple receive antennas with various wireless channel states.

Before resource allocation, the curve-fitting approach is utilized to find the optimal mathematical representation of the relationship between semantic compression ratio and classification accuracy. The fitted parameters and the MSE of each task are given in Table III, which provide guidance for the following experiments. As the compression ratio grows, the number of feature maps actually transmitted may not be enough for the classifier to recognize the attributes of objects in original images. Although the task performance inevitably degrades, a suitable compression ratio can control the size of transmitted data and maintain the satisfactory accuracy.

TABLE III
FITTING PARAMETERS.

Parameter	Task 1 (j=1)	Task 2 (j=2)	Task 3 (j=3)
α_1^j	-0.633	-0.639	-0.737
α_2^j	15.408	19.136	12.474
α_3^j	0.925	0.901	0.917
MSE	9.128×10^{-5}	1.911×10^{-4}	2.269×10^{-4}

In order to implement a long-term resource allocation, the actor and critic learn in different simulation scenarios based on the above parameters and strive to maximize the reward value. To demonstrate the performance gain of the DDPG algorithm, this paper compares the proposed scheme with the following three baseline schemes:

- Asynchronous advantage actor-critic (A3C) driven resource allocation scheme: By creating multiple workers to interact with the environment in parallel, the learned gradients are propagated to a global network. A3C algorithm applies the idea of parallel computing, which improves the utilization of computing resources. Since A3C algorithm can deal with continuous and discrete action spaces, it is widely used in communication resource scheduling.
- The greedy transmission scheme: The goal of the greedy transmission scheme is to maximize short-term gains, namely maximize the quantity of packets successfully received by the edge server within a time slot. This scheme is equivalent to pursuing the maximum system sum rate at the technical level and does not consider semantic compression. It is implemented based on the DDPG framework, and its state space, action space and immediate rewards are consistent with the proposed scheme. In particular, both the semantic compression ratio of UEs and the reward discount factor are set to 0.

- The greedy transmission scheme combined with semantic compression: Similar to the proposed scheme, this scheme jointly optimizes the bandwidth, power and semantic compression ratio of UEs. The purpose of this scheme is to maximize the transmission efficiency of tasks in a slot as much as possible. It is implemented based on the DDPG framework, and its state space, action space and immediate rewards are consistent with the proposed scheme. In particular, the reward discount factor is set to 0.

Fig. 6 demonstrates the convergence of the proposed DDPG-driven resource allocation schemes and the above three baseline schemes with $P_{\max}$=0.2 W, and $B_{\max}$=3 MHz. It can be seen that the reward values of all schemes show a stable convergence trend with the increase of iterations. The proposed scheme can obtain a larger reward, which means a better balance between the accuracy and number of executions of classification tasks. Obviously, the proposed DDPG scheme achieves the highest reward value and relatively fast convergence speed. Although the A3C algorithm requires multiple pairs of actors and critics to explore the best actions, the practice of placing actors and critics in multiple threads for synchronous training greatly reduces the training time. However, the asynchronous learning mode does not lead to higher reward values. Both the greedy transmission schemes with and without semantic compression seek to maximize the benefit within a time slot, and the obtained resource allocation strategy is not optimal in the long run.

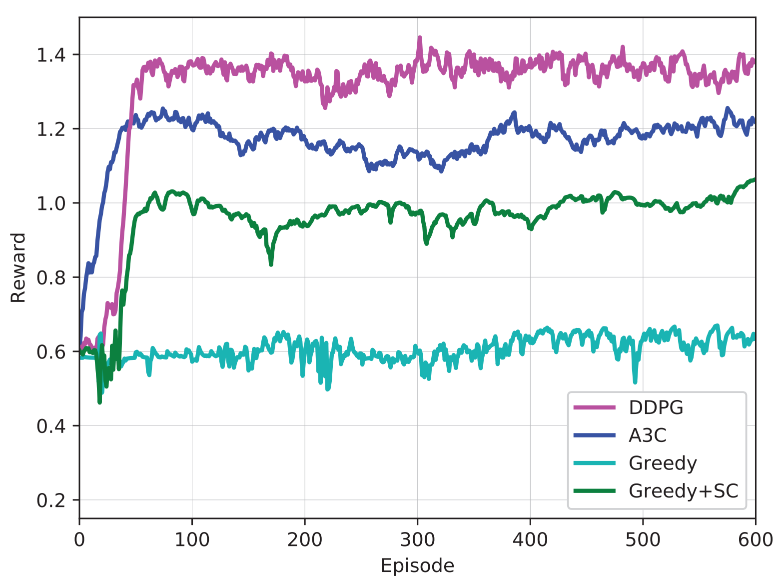


Fig. 6. The rewards for the proposed DDPG scheme and baseline schemes.

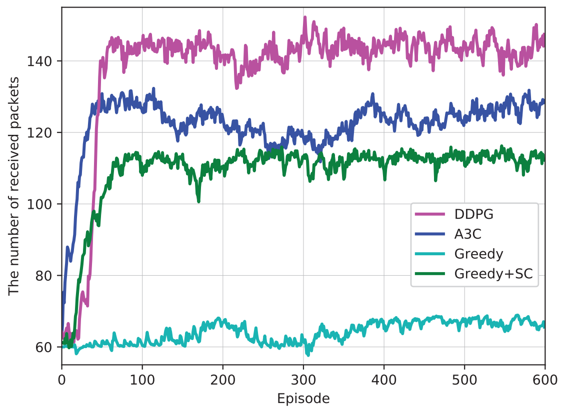


Fig. 7. The number of received packets with increasing iterations.

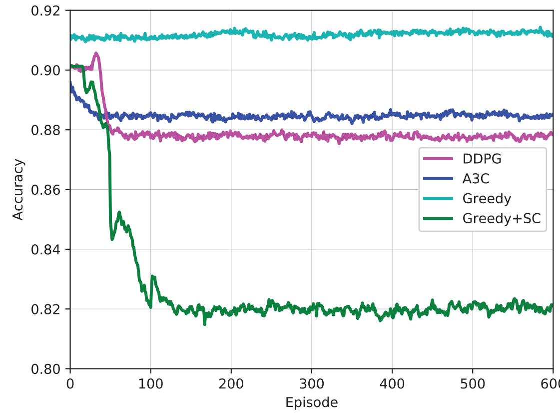


Fig. 8. Classification accuracy with increasing iterations.

A more specific performance of the considered four resource allocation schemes can be observed in Fig. 7 and Fig. 8. The advantage of the proposed resource allocation scheme is further verified from the perspectives of the quantity of packets received by the edge server and the achievable average classification accuracy in a period. Benefiting from the semantic compression and the intelligent decision-making capability of the DDPG algorithm, the proposed resource allocation scheme can significantly increase the quantity of packets received by the edge server with a reasonable loss of task accuracy. In the output action set of the A3C-based scheme, more semantic features are preserved than the DDPG-based scheme, which leads to a reduction in the quantity of classification tasks processed by the receiver in one period. The greedy transmission scheme sends all the extracted semantic features, maintaining the best image classification accuracy. However, when the wireless resources are limited, the greedy transmission scheme may be difficult to process the classification task in a timely manner. The greedy transmission scheme considering semantic compression reduces the size of the data packet so that the data in the buffer of each UE in the current time slot can be sent as much as possible. However, excessive semantic compression will lead to unsatisfactory task accuracy.

The total number of packets received by the edge server versus the maximum available bandwidth is depicted in Fig. 9. For these schemes considering semantic compression, the quantity of packets received by the edge server increases with more available bandwidth and gradually converges. This is because the distinction between maximizing long-term benefits and maximizing short-term benefits will be narrowed when the total available bandwidth is

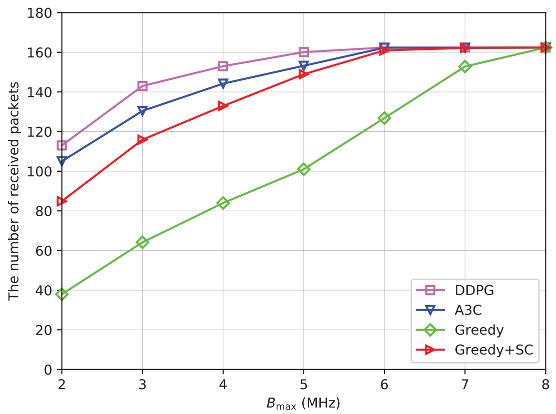


Fig. 9. The number of received packets versus the maximum bandwidth.

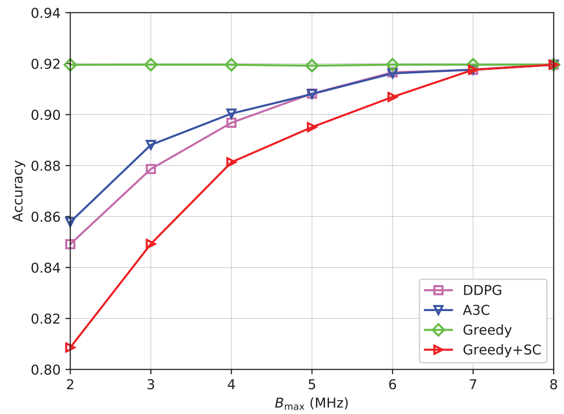


Fig. 10. Task accuracy versus the maximum bandwidth.

relatively sufficient. The greedy transmission scheme sends all feature maps and treats them indiscriminately, thus the latency cost it needs to bear is much higher than the other two schemes, which is fatal to latency-sensitive tasks. In the case of extremely scarce bandwidth, the proposed scheme is more competitive than the other three schemes.

The image classification accuracy achieved by the above four resource allocation schemes with different maximum bandwidths is investigated in Fig.10. Consistent with the prediction, the greedy transmission scheme achieves the best accuracy whether the bandwidth is relatively sufficient or scarce. In terms of task accuracy, the DDPG-based and A3C-based resource allocation schemes are more advantageous than the greedy transmission scheme combined with semantic compression that pursue reward maximization within a single slot. The proposed scheme achieves comparable classification accuracy to the A3C-based scheme. In addition, in the case of limited bandwidth resources, the proposed scheme can transmit 10%-20% more data packets than the A3C-based scheme.

Under the same available bandwidth $B_{\max}$=3 MHz, the quantity of data packets arriving at the receiver achieved by the above four resource allocation schemes with different maximum transmit power is investigated in Fig. 11. When the total transmit power is increased from 140 mW to 200 mW, the greedy transmission scheme has little improvement in the quantity of packets received. All semantic enabled resource allocation schemes outperforms the greedy scheme that pursue the system sum rate at the technical level. It can be observed that the practice of jointly optimizing compression ratio, bandwidth, and transmit power is suitable for machine-to-machine

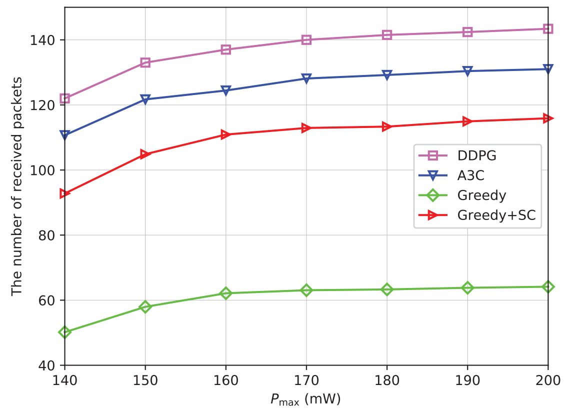


Fig. 11. The number of received packets versus the maximum transmit power.

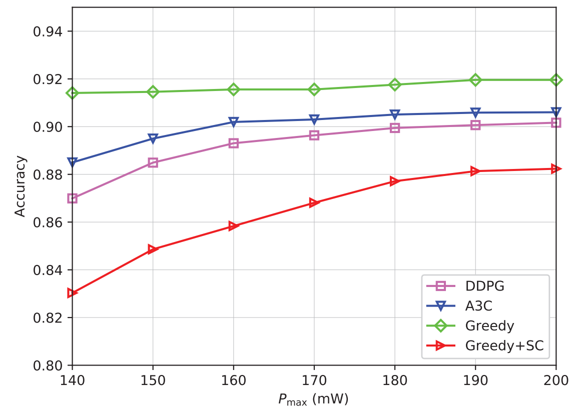


Fig. 12. Task accuracy versus the maximum transmit power.

communications with low power consumption.

Fig. 12 shows the impact of total available power on the task accuracy of the four schemes with the maximum available bandwidth $B_{\max}$=3 MHz. The greedy transmission scheme combined with semantic compression brings a relatively large accuracy penalty to AI tasks. Both the proposed DDPG scheme and A3C scheme approach the upper bound of the classification accuracy in high transmit power regions. Fig. 11 and Fig. 12 again prove that the proposed DDPG-driven resource allocation scheme can sacrifice a reasonable task accuracy in exchange for maximizing the transmission efficiency of tasks. It is meaningful to ensure the execution quality of AI tasks and alleviate communication pressure in scenarios with limited wireless resources.

V. Conclusion

In this paper, a novel DRL-driven resource allocation scheme with the constraints of limited wireless resource for task-oriented semantic communication network was proposed. Different from traditional communication modes that focused on technical-level metrics, the proposed scheme assigned corresponding priority to data based on its contribution to the correct execution of AI tasks, and controlled the amount of data actually transmitted according to currently available wireless resources. Moreover, a joint optimization problem of the semantic feature compression ratio, transmit power, and bandwidth of each intelligent device was formulated to maximize the long-term transmission efficiency of tasks. In order to quickly arrive at the optimal solution of this problem, a DDPG agent was trained in simulated scenarios where intelligent devices have different task assignments to perform dynamic resource management. The experimental results demonstrate that the proposed scheme can significantly increases the number of packets successfully transmitted by users with a reasonable performance penalty in resource-limited wireless networks.

References

[1] K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6G: Vision, enabling technologies, and applications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 5–36, Jan. 2021.

[2] H. Joshi, S. Santra, S. J. Darak, M. K. Hanawal, and S. V. S. Santosh, "Multiplay multiarmed bandit algorithm based sensing of noncontiguous wideband spectrum for AIoT networks," *IEEE Trans. Ind. Inf.*, vol. 18, no. 5, pp. 3337–3348, May 2022.

[3] J. Wang, C. Jiang, H. Zhang, Y. Ren, K.-C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to pareto-optimal wireless networks," *IEEE Commun. Surv. Tutorials*, vol. 22, no. 3, pp. 1472–1514, Jan. 2020.

[4] O. Runsewe and N. Samaan, "Cloud resource scaling for time-bounded and unbounded big data streaming applications," *IEEE Trans. Cloud Comput.*, vol. 9, no. 2, pp. 504–517, Oct. 2021.

[5] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Commun. Mag.*, vol. 59, no. 8, pp. 44–50, Aug. 2021.

[6] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, Apr. 2021.

[7] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.

[8] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cognit. Commun. Networking*, vol. 5, no. 3, pp. 567–579, Sep. 2019.

[9] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142–153, Nov. 2020.

[10] Y. Yang, C. Guo, F. Liu, C. Liu, L. Sun, Q. Sun, and J. Chen, "Semantic communications with artificial intelligence tasks: Reducing bandwidth requirements and improving artificial intelligence task performance," *IEEE Industrial Electronics Magazine*, pp. 2–11, 2022.

[11] C. Bhar and E. Agrell, "Energy- and bandwidth-efficient, QoS-aware edge caching in fog-enhanced radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 9, pp. 2762–2771, Mar. 2021.

[12] J. Wang, W. Cheng, and H. Zhang, "Caching and D2D assisted wireless emergency communications networks with statistical QoS provisioning," *J. Commun. Inf. Networks*, vol. 5, no. 3, pp. 282–293, Sept. 2020.

[13] N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya, "Streaming video QoE modeling and prediction: A long short-term memory approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 661–673, Jan. 2020.

[14] Z. Zhou, Y. Guo, Y. He, X. Zhao, and W. M. Bazzi, "Access control and resource allocation for M2M communications in industrial automation," *IEEE Trans. Ind. Inf.*, vol. 15, no. 5, pp. 3093–3103, Mar. 2019.

[15] S. Dawaliby, A. Bradai, Y. Pousset, and C. Chatellier, "Joint energy and QoS-aware memetic-based scheduling for M2M communications in LTE-M," *IEEE Trans. Emerging Top. Comput. Intell.*, vol. 3, no. 3, pp. 217–229, Dec. 2019.

[16] J. Shao, Y. Mao, and J. Zhang, "Learning task-oriented communication for edge inference: An information bottleneck approach," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 197–211, Nov. 2022.

[17] H. Zhu, P. Tiwari, A. Ghoneim, and M. S. Hossain, "A collaborative AI-enabled pretrained language model for AIoT domain question answering," *IEEE Trans. Ind. Inf.*, vol. 18, no. 5, pp. 3387–3396, May 2022.

[18] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource allocation for text semantic communications," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1394–1398, July 2022.

[19] C. Liu, C. Guo, Y. Yang, and N. Jiang, "Adaptable semantic compression and resource allocation for task-oriented communications," *arXiv preprint arXiv:2204.08910*, Apr. 2022.

[20] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Ye Li, "QoE-aware resource allocation for semantic communication networks," in *GLOBECOM 2022*, Dec. 2022, pp. 3272–3277.

[21] L. Xia, Y. Sun, X. Li, G. Feng, and M. A. Imran, "Wireless resource management in intelligent semantic communication networks," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, May 2022, pp. 1–6.

[22] H. Zhang, N. Yang, W. Huangfu, K. Long, and V. C. M. Leung, "Power control based on deep reinforcement learning for spectrum sharing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4209–4219, Mar. 2020.

[23] Z. Ding, R. Schober, and H. V. Poor, "No-pain no-gain: DRL assisted optimization in energy-constrained CR-NOMA networks," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5917–5932, Jun. 2021.

[24] S. Wang, T. Lv, W. Ni, N. C. Beaulieu, and Y. J. Guo, "Joint resource management for MC-NOMA: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5672–5688, Sep. 2021.

[25] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, Jul. 2019.

[26] Y. Wang, W. Fang, Y. Ding, and N. Xiong, "Computation offloading optimization for UAV-assisted mobile edge computing: A deep deterministic policy gradient approach," *Wireless Networks*, vol. 27, no. 4, pp. 2991–3006, May 2021.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[28] R. Fan, H. Wang, Y. Wang, M. Liu, and I. Pitas, "Graph attention layer evolves semantic segmentation for road pothole detection: A benchmark and algorithms," *IEEE Trans. Image Process.*, vol. 30, pp. 8144–8154, Sept. 2021.

[29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[30] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, "Long-tail learning via logit adjustment," *arXiv preprint arXiv:2007.07314*, Jul. 2020.

[31] Z. Shi, X. Xie, H. Lu, H. Yang, J. Cai, and Z. Ding, "Deep reinforcement learning-based multidimensional resource management for energy harvesting cognitive NOMA communications," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3110–3125, May 2022.

[32] F. A. Galatolo, M. G. Cimino, and G. Vaglini, "Solving the scalarization issues of advantage-based reinforcement learning algorithms," *Comput. Electr. Eng.*, vol. 92, pp. 107–117, Jun. 2021.