



# Cooperative and Collaborative Multi-Task Semantic Communication for Distributed Sources

Ahmad Halimi Razlighi , Maximilian H. V. Tillmann ,  
Edgar Beck , Carsten Bockelmann , and Armin Dekorsy 

Department of Communications Engineering, University of Bremen, Germany  
E-mails: {halimi, tillmann, beck, bockelmann, dekorsy}@ant.uni-bremen.de

**Abstract**—In this paper, we explore a multi-task semantic communication (SemCom) system for distributed sources, extending the existing focus on collaborative single-task execution. We build on the cooperative multi-task processing introduced in [1], which divides the encoder into a common unit (CU) and multiple specific units (SUs). While earlier studies in multi-task SemCom focused on full observation settings, our research explores a more realistic case where only distributed partial observations are available, such as in a production line monitored by multiple sensing nodes. To address this, we propose an SemCom system that supports multi-task processing through cooperation on the transmitter side via split structure and collaboration on the receiver side. We have used an information-theoretic perspective with variational approximations for our end-to-end data-driven approach. Simulation results demonstrate that the proposed cooperative and collaborative multi-task (CCMT) SemCom system significantly improves task execution accuracy, particularly in complex datasets, if the noise introduced from the communication channel is not limiting the task performance too much. Our findings contribute to a SemCom framework capable of handling distributed sources and multiple tasks simultaneously, advancing the applicability of SemCom systems in real-world scenarios.

**Index Terms**—Semantic communication, cooperation, collaboration, multi-tasking, infomax, deep learning.

## I. INTRODUCTION

Recent breakthroughs in artificial intelligence, particularly in deep learning (DL) and end-to-end (E2E) communication technologies, have led to the rise of semantic communication (SemCom) [2]. It has attracted significant attention, being recognized as a critical enabler for the sixth generation (6G) of wireless communication networks. SemCom is expected to play a key role in supporting a wide range of innovative applications that will define 6G connectivity and beyond [3].

In contrast to conventional communication systems, which are designed based on Shannon's information theory and focus on the accurate transmission of symbols, SemCom prioritizes understanding the meaning and goals behind transmitted information. SemCom operates at the second level of communication, the semantic level, where the goal is to convey the desired meaning rather than ensuring exact bit-level accuracy [4]. By surpassing the traditional focus on the precise transmission of bits, SemCom is well-suited for emerging

This work was supported in part by the German Ministry of Education and Research (BMBF) under Grant 16KISK016 (Open6GGub) and the German Research Foundation (DFG) under grant 500260669 (SCIL).

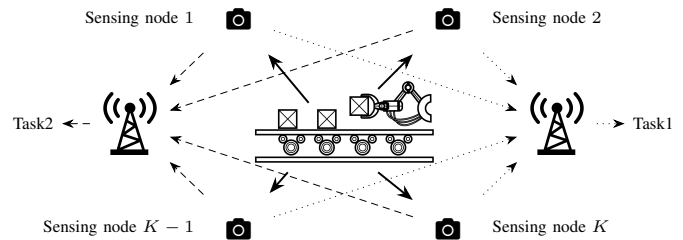


Fig. 1: An example of cooperative and collaborative multi-tasking for distributed sources.

applications, such as the industrial internet and autonomous systems, where successful task execution is prioritized over the exact reconstruction of transmitted data at the receiver.

Research into SemCom has explored five main approaches, with four detailed in [5] and a fifth inspired by Weaver's extension of Shannon's theory to include the semantic level [6]. These approaches are:

- Classical approach: Quantifies semantic information using logical probability.
- Knowledge graph (KG) approach: Represents semantics through structured KGs.
- Machine learning (ML) approach: Encodes semantics within learned model parameters.
- Significance approach: Focuses on timing as a key component of semantic meaning.
- Information theory approach: Extends Shannon's framework to address semantic-level communication.

Recent works in SemCom primarily focus on two research directions: data reconstruction and task execution. Initial investigations into data recovery were led by [7] and [8], which utilized ML techniques to reconstruct diverse data sources such as text, speech, and images. Building on these foundational works, [9] and [10] have extended the focus to explore concepts like communication efficiency in SemCom. In addition, systems dealing with structured data have been examined through the KG approach to enhance data recovery [11].

In task-oriented SemCom, the focus shifts to executing intelligent tasks at the receivers. Most research in this area

has concentrated on single-task scenarios. For example, [12] developed a communication scheme using the information bottleneck framework, which encodes information while adapting to dynamic channel conditions.

Moreover, some works considered more realistic scenarios in which the source is distributed, for instance, [13] studied distributed relevant information encoding for collaborative feature extraction to fulfill a task. [14] also offered a framework for collaborative retrieval of the message using multiple received semantic information.

To address practical communication scenarios, SemCom systems must be capable of handling multiple tasks simultaneously. Early efforts, such as [15] and [16], explored non-cooperative methods where each task operates on its respective dataset independently. Conversely, recent works like [17], [18], and [19] studied joint multi-tasking using established ML approaches and architecture.

The prior works on multi-tasking in SemCom have focused on ML approaches, however, [1] introduced an information-theoretic analysis of the problem. This study proposed a split structure for the semantic encoder, dividing the semantic encoder into a common unit (CU) and multiple specific units (SUs), to enable cooperative processing of various tasks. The split structure can perform multiple tasks based on a single observation. In this work, we have examined a more applicable scenario, in which the full observation is not accessible, however, we have different distributed views of our main observation. This is illustrated in Fig. 1, where the full observation is the whole view of the production line but our sensing nodes provide partial observations from different positions.

In distributed cases, where the source/observation is distributed, the task cannot be executed depending on a single sensing node alone. Thus, collaborative systems, where nodes jointly complete a single task at the receiver side, have been studied in [14] and [13]. However, our research expands the multi-task scenarios to distributed settings, by combining the split structure from [1], which brings cooperation of multiple tasks on the transmitter side, with collaboration on the receiver side. Therefore, we have contributed to a SemCom system capable of cooperatively and collaboratively executing multiple tasks by proposing a cooperative and collaborative multi-task (CCMT) SemCom architecture. In our data-driven approach, we have tailored semantic communication to multi-task processing for distributed sources and formulated it through information theory using variational approximations. Key contributions include:

- Combining the cooperative multi-task process, enabled by the split structure, with the collaborative process of the distributed observations.
- Considering different training methods and channel conditions for the proposed CCMT architecture.
- Demonstrating the effectiveness of the CCMT system by showing enhancements in task execution over various scenarios, specifically when dealing with more challenging datasets.

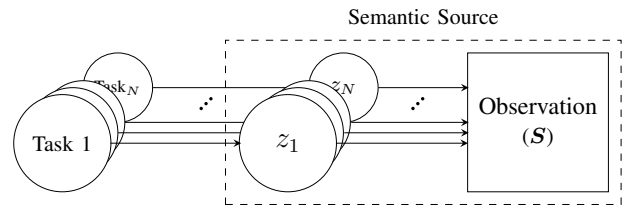


Fig. 2: Probabilistic graphical modeling of the proposed semantic source [1].

## II. SYSTEM MODEL

This section introduces our probabilistic modeling of the proposed CCMT model. Furthermore, we formulate an information-theoretic optimization problem that aims to optimize the joint execution of multiple tasks in an E2E manner.

### A. System Probabilistic Modeling

Fig. 2 illustrates our interpretation of *semantic source* as discussed in [1]. Such a definition enables the simultaneous extraction of multiple *semantic variables* based on a single observation and addresses multiple tasks. It consists of  $N$  semantic variables, denoted by  $\mathbf{z} = [z_1 z_2 \dots z_N]$ , lying behind an observation  $\mathcal{S}$ , and the given tasks specify one/multiple semantics to be of our interest. The tuple of  $(\mathbf{z}, \mathcal{S})$  is defined as the semantic source, fully described by the probability distribution of  $p(\mathbf{z}, \mathcal{S})$ . In this study, we focus on a more practical scenario of the distributed setting, where instead of the full observation, multiple partial views of the data are available. These partial observations, denoted as  $\mathcal{S}_1, \dots, \mathcal{S}_K$ , each contain information about some or all semantic variables.

In [1], it was demonstrated that when semantic variables share statistical relationships, a split semantic encoder, comprising a CU and multiple SUs, enables cooperative SemCom, significantly improving performance in multi-task cases by utilizing common information. In realistic scenarios, sensing nodes only access partial observations of the source, and collaborative approaches are required to perform tasks based on these distributed views.

To integrate cooperative multi-tasking with collaborative handling of distributed data, our proposed system model consists of  $K$  observations available at  $K$  sensing nodes and  $N$  semantic variables, each associated with a unique task. As illustrated in Fig. 3, at each sensing node, the CU encoder first extracts the common relevant information from its observation. Next,  $N \times K$  SU encoders extract and transmit task-specific information to their respective decoders for collaborative decoding. To fulfill each task, the corresponding receivers need  $K$  transmitted information extracted and transmitted by assigned SUs from each observation. Since we consider the execution of two tasks, two SUs are required at each sensing node. Thus, we show the output of SU encoders as  $\mathbf{x}_{1,1}, \mathbf{x}_{1,2}, \dots, \mathbf{x}_{K,1}, \mathbf{x}_{K,2}$ , and their noise-corrupted version received at the corresponding decoders are indicated by  $\hat{\mathbf{x}}_{1,1}, \hat{\mathbf{x}}_{1,2}, \dots, \hat{\mathbf{x}}_{K,1}, \hat{\mathbf{x}}_{K,1}$ .

Our approach incorporates wireless channels between encoders and decoders, employing the additive white Gaussian

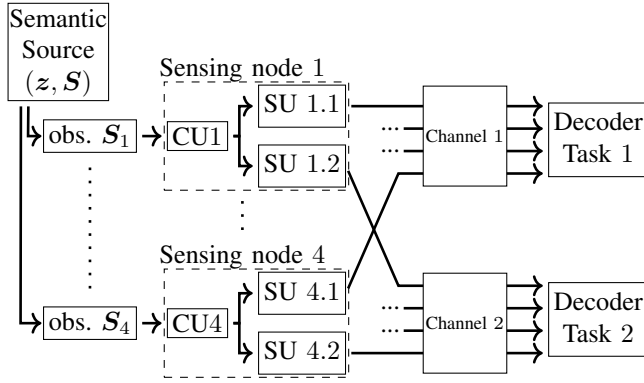


Fig. 3: An illustration of the proposed CCMT system model for distributed partial observations for  $N=2$  and  $K=4$ .

noise (AWGN) channel. As shown generally in Fig. 4, the Markov representation of our system model for the  $i$ -th semantic variable is outlined for  $\forall k \in \{1, \dots, K\}$  as follows.

$$p(\hat{z}_i, \hat{\mathbf{x}}_{k,i}, \mathbf{x}_{k,i}, \mathbf{c}_k | \mathbf{S}_k) = p^{\text{Dec}_i}(\hat{z}_i | \hat{\mathbf{x}}_{k,i}) p^{\text{Channel}_k}(\hat{\mathbf{x}}_{k,i} | \mathbf{x}_{k,i}) p^{\text{SU}_{k,i}}(\mathbf{x}_{k,i} | \mathbf{c}_k) p^{\text{CU}_k}(\mathbf{c}_k | \mathbf{S}_k). \quad (1)$$

In (1),  $p^{\text{CU}_k}(\mathbf{c}_k | \mathbf{S}_k)$  defines the CU of the  $k$ -th sensing node, which extracts the common relevant information available in the  $k$ -th observation amongst all tasks. The corresponding SU for  $i$ -th semantic variable at  $k$ -th sensing node is described by  $p^{\text{SU}_{k,i}}(\mathbf{x}_{k,i} | \mathbf{c}_k)$  extracting task-specific information and providing  $\mathbf{x}_{k,i}$  as the channel input. The corresponding decoder is then specified by  $p^{\text{Dec}_i}(\hat{z}_i | \hat{\mathbf{x}}_{k,i})$ , where  $\hat{\mathbf{x}}_{k,i} \in \mathbb{R}^{m_i}$ ,  $\forall k \in \{1, \dots, K\}$  is the received information passed through the AWGN channel and modeled like  $\hat{\mathbf{x}}_{k,i} = \mathbf{x}_{k,i} + \mathbf{n}$ , where  $\mathbf{n} \sim \mathcal{N}(\mathbf{0}_{m_i}, \sigma_i^2 \mathbf{I}_{m_i})$ , and  $m_i$  is the size of the encoded task-specific information or the number of channel uses.

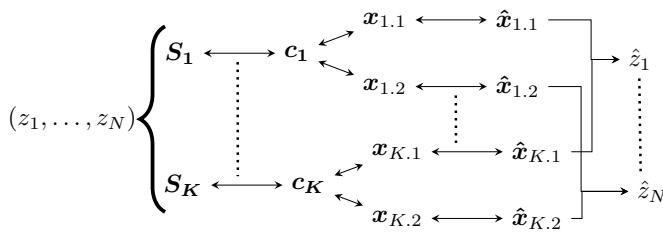


Fig. 4: Markov representation of the CCMT system.

### B. Optimization Problem

We formulate an optimization problem by adopting the information maximization principle together with the E2E learning method, as follows.

$$\begin{aligned} & [p^{\text{CU}_k}(\mathbf{c}_k | \mathbf{S}_k)^*, p^{\text{SU}_k}(\mathbf{x}_k | \mathbf{c}_k)^*]_{k=1}^K = \\ & \arg \max_{\substack{p^{\text{CU}_k}(\mathbf{c}_k | \mathbf{S}_k), \\ p^{\text{SU}_k}(\mathbf{x}_k | \mathbf{c}_k)}} \sum_{i=1}^N I(z_i; \hat{\mathbf{x}}_{(1:K),i}). \quad (2) \end{aligned}$$

Thus, the objective is to maximize the mutual information between the channel outputs  $\hat{\mathbf{x}}_{(1:K),i}$  of the corresponding SUs, and the semantic variables  $z_i$ . Expanding the mutual information in (2) as discussed in detail in [1], the approximated objective function is derived like:

$$\begin{aligned} \mathcal{L}^{\text{CCMT}}(\boldsymbol{\theta}, \boldsymbol{\Phi}) &= \sum_{i=1}^N I(z_i; \hat{\mathbf{x}}_{(1:K),i}) \\ &\approx \mathbb{E}_{p_{\boldsymbol{\theta}}^{\text{CU}}(\mathbf{c}_{1:K} | \mathbf{S}_{1:K})} \left[ \sum_{i=1}^N \left\{ \mathbb{E}_{p(\mathbf{S}_{1:K}, z_i)} \left[ \mathbb{E}_{p_{\boldsymbol{\Phi}}^{\text{SU}}(\hat{\mathbf{x}}_{(1:K),i} | \mathbf{c}_{1:K})} [\log p(z_i | \hat{\mathbf{x}}_{(1:K),i})] \right] \right\} \right]. \quad (3) \end{aligned}$$

To derive the objective function on (3), we have employed the variational method, which is a way to approximate intractable computations based on some adjustable parameters, like weights in NNs [20]. The technique is widely used in machine learning, e.g., [21], and also in task-oriented communications, e.g., [12], [13], and [14]. Thus, our posterior distributions,  $\{p^{\text{CU}_k}(\mathbf{c}_k | \mathbf{S}_k)\}_{k=1}^K$  and  $\{p^{\text{SU}_k}(\hat{\mathbf{x}}_k | \mathbf{c}_k)\}_{k=1}^K$  are approximated by NN parameters of  $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^K$  and  $\boldsymbol{\Phi} = \{\phi_k\}_{k=1}^K$  respectively.

As shown in (3), by considering the channel outputs we aim to emphasize the role of joint semantic and channel coding performed by our SUs. Employing the fact that  $p_{\phi_k}^{\text{SU}_k}(\hat{\mathbf{x}}_k | \mathbf{c}_k) = \int p_{\phi_k}^{\text{SU}_k}(\mathbf{x}_k | \mathbf{c}_k) p^{\text{Channel}_k}(\hat{\mathbf{x}}_k | \mathbf{x}_k) d\mathbf{x}_k$ , we try to optimize  $p_{\phi_k}^{\text{SU}_k}(\hat{\mathbf{x}}_k | \mathbf{c}_k)$ .

Regarding the  $i$ -th decoder in (3), the  $p^{\text{Dec}_i}(z_i | \hat{\mathbf{x}}_{(1:K),i})$  can be fully determined using the known distributions and underlying probabilistic relationship in (1) as:

$$p^{\text{Dec}_i}(z_i | \hat{\mathbf{x}}_{k,i}) = \frac{\int p_{\phi_k}^{\text{SU}_k}(\hat{\mathbf{x}}_{k,i} | \mathbf{c}_k) p_{\theta_k}^{\text{CU}_k}(\mathbf{c}_k | \mathbf{S}_k) p(\mathbf{S}_k, z_i) d\mathbf{s}_k d\mathbf{c}_k}{p(\hat{\mathbf{x}}_{k,i})}, \quad (4)$$

however, due to the high-dimensional integrals, (4) becomes intractable and we need to follow the variational approximation technique, resulting in the following:

$$\begin{aligned} \mathcal{L}_{\text{approx}}^{\text{CCMT}}(\boldsymbol{\theta}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) &= \sum_{i=1}^N I(z_i; \hat{\mathbf{x}}_{(1:K),i}) \\ &\approx \mathbb{E}_{p_{\boldsymbol{\theta}}^{\text{CU}}(\mathbf{c}_{1:K} | \mathbf{S}_{1:K})} \left[ \sum_{i=1}^N \left\{ \mathbb{E}_{p(\mathbf{S}_{1:K}, z_i)} \left[ \mathbb{E}_{p_{\boldsymbol{\Phi}}^{\text{SU}}(\hat{\mathbf{x}}_{(1:K),i} | \mathbf{c}_{1:K})} [\log p_{\boldsymbol{\Psi}}^{\text{Dec}_i}(z_i | \hat{\mathbf{x}}_{(1:K),i})] \right] \right\} \right]. \quad (5) \end{aligned}$$

Where in (5),  $\boldsymbol{\Psi} = \{\psi_i\}_{i=1}^N$  represents NN parameters approximating the true distribution of decoders. To obtain the empirical estimate of the above objective function, we approximate the expectations using Monte Carlo sampling assuming the existence of a dataset  $\{\mathbf{S}^{(j)}, z_1^{(j)}, \dots, z_N^{(j)}\}_{j=1}^J$ .

where  $J$  represents the batch size of the dataset [22].

$$\mathcal{L}_{\text{empir.}}^{\text{CCMT}}(\boldsymbol{\theta}, \Phi, \Psi) \approx \frac{1}{L} \sum_{l=1}^L \left[ \sum_{i=1}^N \left\{ \frac{1}{J} \sum_{j=1}^J \left[ \frac{1}{T} \sum_{t=1}^T [\log q_{\psi_i}^{\text{Dec}_i}(z_i | \hat{\mathbf{x}}_{(1:K),j,t})] \right] \right\} \right]. \quad (6)$$

In (6),  $L$  represents the sample size of the cooperative processing, and  $T$  is the channel sampling size for each batch.

### III. SIMULATION RESULTS

For the distributed scenario, we compare our proposed CCMT with a baseline approach called single-task collaborative (STC) semantic communication with no cooperative multitask processing. We evaluate the task execution error rate for task 1 and task 2 across different training scenarios, channel conditions, and NN sizes, demonstrating the performance improvements of CCMT over STC<sup>1</sup>.

#### A. Simulation Setup

We evaluate our proposed architecture for four sensing nodes that need to collaborate and cooperate on two tasks, where each sensing node has a partial observation of exactly one quarter of the full image showing a digit of the MNIST dataset [23]. For our simulations, we consider task 1 to be the binary classification of digit two and task 2, the categorical classification of the digits in the MNIST dataset. To examine more practical and complex scenarios, in which sensing nodes may view the observing object from different angles, we consider situations where each quarter image is individually rotated by for a randomly selected angle.

For the semantic encoders we use convolutional NNs (CNNs) with ReLU activation functions and max-pool layers for image size reduction, as specified in Tab. I. In our simulations, we fix the CU to two CNN layers and the SU to one CNN layer with a fully connected (FC) layer for each task. In the STC case, where no CU is used, the SUs consist of three CNN layers and one FC layer each, to make the number of layers equivalent to the CCMT case for a fair comparison. The number of filters for each CNN layer in CCMT:  $c_1, c_2, c_3$  and STC:  $k_1, k_2, k_3$ , is set for each comparison such that the total number of parameters in CCMT and STC are approximately equal.

For all simulations, the number of channel uses for task 1 and task 2 is two per sensing node, i.e.,  $m_1 = m_2 = 2$ . For Figs. 5 and 6,  $c_1 = 6, c_2 = 5, c_3 = 3$ , and  $k_1 = 4, k_2 = 4, k_3 = 3$  are used, which means for the CNN layers of the CCMT architecture a total of  $(9+1)c_1 + (9c_1+1)c_2 + 2((9c_2+1)c_3) = 611$  parameters and for the STC architecture 598 parameters are used. The final layer of each encoder normalizes the output power across the channel uses to average the output power for each transmission to one over each NN training batch. The signal-to-noise ratio (SNR) of the AWGN channel is defined as  $\text{SNR} := 1/\sigma_i^2$ , where  $\sigma_i^2$  is the noise

TABLE I: The NN structure for each distributed sensing node

	Output size	No. of param.
<b>Encoder Layers for CCMT (for task 1 and 2)</b>		
CU: CNN layer, ReLU, max-pool	$7 \times 7 \times c_1$	$(9+1)c_1$
CU: CNN layer, ReLU, max-pool	$3 \times 3 \times c_2$	$(9c_1+1)c_2$
SU <sub>task1</sub> : CNN layer, ReLU	$3 \times 3 \times c_3$	$(9c_2+1)c_3$
SU <sub>task1</sub> : FC, power normalization	$m_1$	$(9c_3+1)m_1$
SU <sub>task2</sub> : CNN layer, ReLU	$3 \times 3 \times c_3$	$(9c_2+1)c_3$
SU <sub>task2</sub> : FC, power normalization	$m_2$	$(9c_3+1)m_2$
<b>Encoder Layers for STC (for task <math>i = 1, 2</math>)</b>		
SU <sub>task<i>i</i></sub> : CNN layer, ReLU, max-pool	$7 \times 7 \times k_1$	$(9+1)k_1$
SU <sub>task<i>i</i></sub> : CNN layer, ReLU, max-pool	$3 \times 3 \times k_2$	$(9k_1+1)k_2$
SU <sub>task<i>i</i></sub> : CNN layer, ReLU,	$3 \times 3 \times k_3$	$(9k_2+1)k_3$
SU <sub>task<i>i</i></sub> : FC, power normalization	$m_i$	$(9k_3+1)m_i$
<b>Decoder</b>		
Decoder task 1: FC, Tanh	16	$(m_1+1)16$
Decoder task 1: FC, Sigmoid	1	$16+1$
Decoder task 2: FC, Tanh	16	$(m_2+1)16$
Decoder task 2: FC, Softmax	10	$(16+1)10$

power of the zero mean i.i.d. Gaussian noise vector of each channel  $\mathbf{n} \in \mathbb{R}^{m_i}$ . For the simulations, 60 000 training and 10 000 validation data samples are used, the results are shown for the validation dataset, and the results of all simulations are averaged over 25 independent iterations.

#### B. Training Scenarios

In Fig. 5 the task execution error rate of task 1 and task 2 over the number of epochs for different SNRs is shown. It is worth mentioning that the SNR is uniformly distributed for all ranges, and specifically for Fig. 5, the evaluation SNR range is the same as the training SNR range. In this figure, the proposed CCMT architecture is compared to the STC architecture when both are trained for 500 epochs for an SNR range from 9 to 11 dB. It can be seen that the CCMT outperforms the STC in task execution error rate.

To further investigate the joint semantic and channel coding performance of the SUs to deal with different channel conditions, the CCMT is first trained for a wide SNR range from  $-10$  to  $20$  dB for 250 epochs, causing the CU to be generalized and then the CU is frozen and the SUs are further trained for the smaller target SNR range of 9 to 11 dB for additional 250 epochs. It can be seen in Fig. 5 that this approach, named ‘‘CCMT-generalized-CU’’, performs equally, or even slightly better than the case where the whole CCMT is trained for a small SNR range. The validation SNR for all results in Fig. 5 is set to 10 dB.

We conclude that the generalization of the CU has the advantage that only retraining of the SUs is required to deal with different channel conditions for task 1 and task 2. Therefore, we use the CCMT-generalized-CU, where the CU is trained for 250 epochs for an SNR range of  $-10$  to  $20$  dB, and

<sup>1</sup>The code is available at <https://github.com/ahmadhalimi95/CCMT>

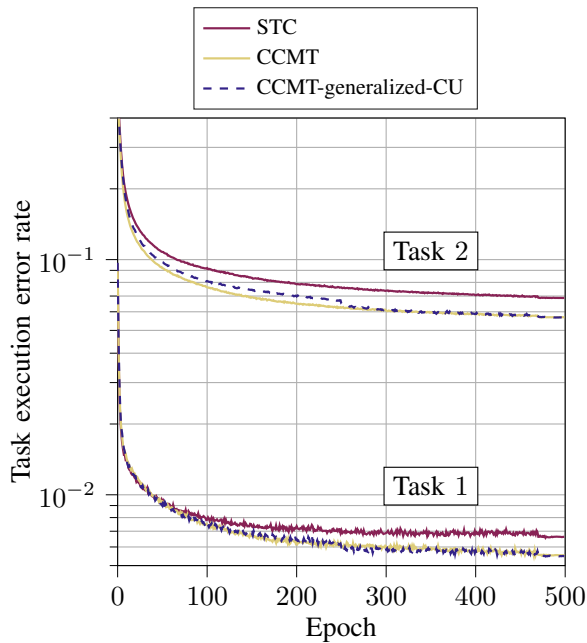


Fig. 5: Task execution error rate over the number of training epochs of task 1 and task 2 for different training scenarios.

then the SUs are trained for specific SNR ranges depending on the channel conditions.

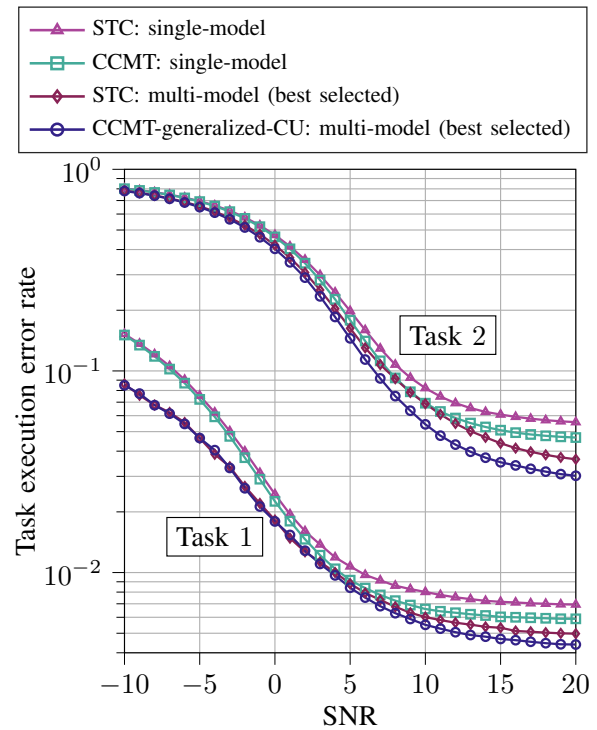
### C. Impact of Different Channel Conditions

Further, the CCMT and STC are compared for different SNR values, in Fig. 6a without, and in Fig. 6b with image rotation, where each partial observation is individually rotated for an angle uniformly distributed between  $\pm 30^\circ$ . For the simulation, multiple SUs models are trained for specific SNR ranges for the CCMT-generalized-CU and the STC, where in the evaluation process the best model is selected for each SNR. These are indicated by “multi-model (best selected)” for both cases in Fig. 6. In total 11 models are trained, with SNR ranges in dB:  $[-12, -10]$ ,  $[-9, -7]$ ,  $[-6, -4]$ ,  $[-3, -1]$ ,  $[0, 2]$ ,  $[3, 5]$ ,  $[6, 8]$ ,  $[9, 11]$ ,  $[12, 14]$ ,  $[15, 17]$ , and  $[18, 20]$ .

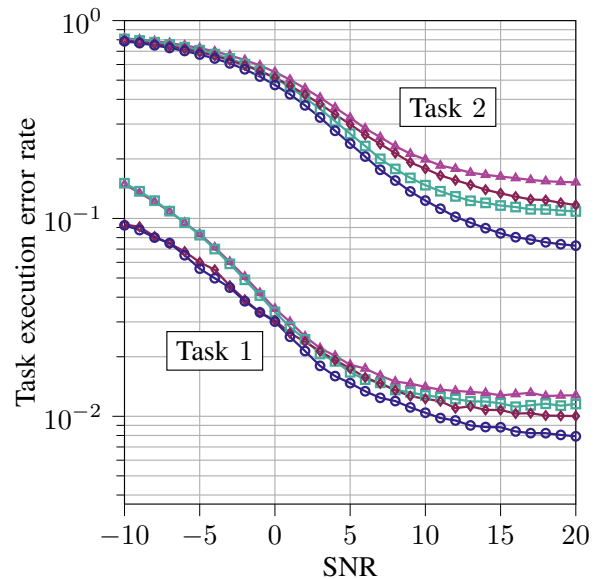
Moreover, the multi-model cases are compared with the “single-model” cases, where only a single model is trained for the SNR range of  $-10$  to  $20$  dB for the CCMT and the STC.

Fig. 6a shows that the multi-model cases (CCMT-generalized-CU and STC) outperform the single-model cases (CCMT and STC) in the whole SNR range for task 1 and for higher SNR values for task 2. Next, the CCMT outperforms the STC for both the single and multi-model cases for higher SNR values. However, for lower SNR values, the CCMT and STC achieve almost the same task execution error rate, as the gain in performance from the CU in cooperative task processing is minimal compared to the errors introduced from the poor channel conditions.

Compared to Fig. 6a, Fig. 6b shows larger gaps in task error rate between the CCMT and STC architectures for both



(a)



(b)

Fig. 6: Task execution error rate over the SNR (a) without image rotation, and (b) with image rotation.

tasks in both single and multi-model cases. This indicates that the CU’s cooperative processing of multiple tasks is more advantageous for more challenging datasets.

### D. Impact of Different NN Sizes

Finally, the CCMT-generalized-CU and STC architectures are compared for different numbers of NN parameters. The

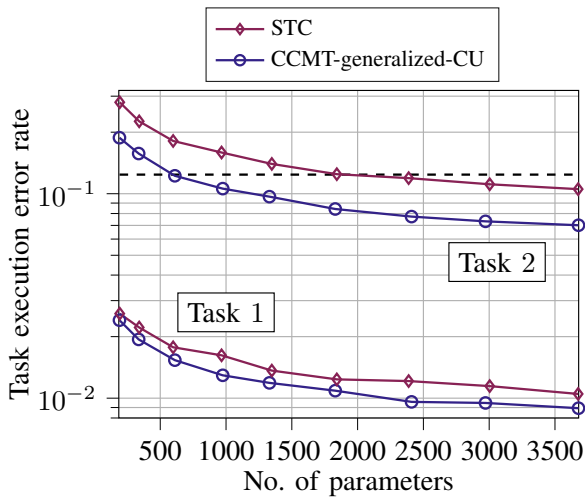


Fig. 7: Task execution error rate over the number of trainable CNN layer parameters. For task 1 the SNR is 5 dB and for task 2 the SNR is 10 dB.

task execution error rates are shown in Fig. 7 over the number of parameters of the CNN layers of each sensing unit for the dataset with rotation.

For this simulation, we consider different channel conditions for task 1 and task 2. For task 1 the validation SNR is 5 dB with an SNR training range from 4 to 6 dB and for task 2 the validation SNR is 10 dB with an SNR training range from 9 to 11 dB. The number of convolution filters is increased from  $c_1 = 4, c_2 = 2, c_3 = 2, k_1 = 2, k_2 = 2, k_3 = 2$  resulting in 190 and 192 parameters, to  $c_1 = 14, c_2 = 13, c_3 = 8, k_1 = 11, k_2 = 10, k_3 = 8$  resulting in 3679 and 3676 parameters, for the CCMT and the STC, respectively. The number of convolution filters in the final CNN layer  $c_3$  and  $k_3$  are always the same for the CCMT and STC. We note that CCMT and STC are trained for 500 epochs in total for each case.

It can be seen that the CCMT saves a significant amount of computing resources compared to STC. For example, for task 2, an error rate of about 0.12 for the STC requires about 1900 NN parameters, while the CCMT requires only about 611. This is illustrated by the dashed line in Fig. 7. Moreover, we observe that in general, increasing the number of parameters decreases the error rate for all cases and the gap between the CCMT and STC stays relatively constant for the investigated parameter range.

#### IV. CONCLUSION

We introduced the CCMT architecture based on an information-theoretic perspective, combining the cooperative processing of multiple tasks with the collaborative processing of distributed observations. We considered different training methods and channel conditions. Simulation results showed that the proposed CCMT architecture lowers the task execution error rate compared to the STC approach, specifically, for more challenging datasets and better channel conditions. Finally, it

was shown that the advantage of the CCMT holds for different NN sizes.

#### REFERENCES

- [1] A. Halimi Razlighi, C. Bockelmann, and A. Dekorsy, "Semantic communication for cooperative multi-task processing over wireless networks," *IEEE Wireless Communications Letters*, vol. 13, no. 10, pp. 2867–2871, 2024.
- [2] "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, pp. 5–41, 11 2022.
- [3] W. Tong and G. Y. Li, "Nine challenges in artificial intelligence and wireless communications for 6g," *IEEE Wireless Communications*, vol. 29, no. 4, pp. 140–145, 2022.
- [4] M. Sana and E. C. Strinati, "Learning semantics: An opportunity for effective 6G communications," Institute of Electrical and Electronics Engineers Inc., 2022, pp. 631–636.
- [5] D. Wheeler and B. Natarajan, "Engineering semantic communication: A survey," *IEEE Access*, vol. 11, pp. 13 965–13 995, 2023.
- [6] W. Weaver, "Recent contributions to the mathematical theory of communication," *ETC: a review of general semantics*, pp. 261–281, 1953.
- [7] H. Xie, Z. Qin, G. Y. Li, and B. H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.
- [8] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 39, pp. 142–153, 1 2021.
- [9] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource allocation for text semantic communications," *IEEE Wireless Communications Letters*, vol. 11, pp. 1394–1398, 7 2022.
- [10] H. Tong, Z. Yang, S. Wang, Y. Hu, W. Saad, and C. Yin, "Federated learning based audio semantic communication over wireless networks," Institute of Electrical and Electronics Engineers Inc., 2021.
- [11] Y. Wang, M. Chen, W. Saad, T. Luo, S. Cui, and H. V. Poor, "Performance optimization for semantic communications: An attention-based learning approach," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.
- [12] J. Shao, Y. Mao, and J. Zhang, "Learning task-oriented communication for edge inference: An information bottleneck approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 197–211, 2022.
- [13] J. Shao, Y. Mao, and J. Zhang, "Task-oriented communication for multidevice cooperative edge inference," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 73–87, 2023.
- [14] E. Beck, C. Bockelmann, and A. Dekorsy, "Semantic information recovery in wireless networks," *Sensors*, vol. 23, p. 6347, 7 2023. [Online]. Available: <https://www.mdpi.com/1424-8220/23/14/6347>
- [15] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2584–2597, 2022.
- [16] G. He, S. Cui, Y. Dai, and T. Jiang, "Learning task-oriented channel allocation for multi-agent communication," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12 016–12 029, 2022.
- [17] Y. Sheng, F. Li, L. Liang, and S. Jin, "A multi-task semantic communication system for natural language processing," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, 2022, pp. 1–5.
- [18] Y. E. Sagduyu, T. Erpek, A. Yener, and S. Ulukus, "Multi - receiver task-oriented communications via multi - task deep learning," in *2023 IEEE Future Networks World Forum (FNWF)*, 2023, pp. 1–6.
- [19] M. Gong, S. Wang, and S. Bi, "A scalable multi-device semantic communication system for multi-task execution," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023, pp. 2227–2232.
- [20] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [21] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, "Deep variational information bottleneck," *arXiv preprint arXiv:1612.00410*, 2016.
- [22] C. M. Bishop, "Pattern recognition and machine learning," *Springer google schola*, vol. 2, pp. 1122–1128, 2006.
- [23] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.