Joint Sensing and Task-Oriented Communications with Image and Wireless Data Modalities for Dynamic Spectrum Access

Yalin E. Sagduyu[1,2], Tugba Erpek[1], Aylin Yener[3], and Sennur Ulukus[4]

[1]Virginia Tech, Arlington, VA, USA
[2]Nexcepta, Gaithersburg, MD, USA
[3]The Ohio State University, Columbus, OH, USA
[4]University of Maryland, College Park, MD, USA

***Abstract*—This paper introduces a deep learning approach leveraging the synergy of multi-modal image and spectrum data for the identification of potential transmitters to assist with dynamic spectrum access. We consider an edge device equipped with a camera that is taking images of potential objects such as vehicles that may harbor transmitters. Recognizing the computational constraints and trust issues associated with on-device computation, we propose a collaborative system wherein the edge device communicates selectively processed information to a trusted receiver acting as a fusion center, where a decision is made to identify whether a potential transmitter is present, or not. We employ task-oriented communications (TOC), utilizing an encoder at the transmitter for joint source coding, channel coding, and modulation. This architecture efficiently transmits essential information of reduced dimension for object classification. Simultaneously, the transmitted signals may reflect off objects and return to the transmitter, allowing for the collection of target sensing data. Then the collected sensing data undergoes a second round of encoding at the transmitter, with the reduced-dimensional information communicated back to the fusion center through TOC. On the receiver side, a decoder performs the task of identifying a transmitter by fusing data samples received through joint sensing and TOC. The two encoders at the transmitter and the decoder at the receiver are jointly trained, enabling a seamless integration of image classification and wireless signal detection. Using AWGN and Rayleigh channel models, we demonstrate the effectiveness of the proposed approach, showcasing high accuracy in transmitter identification across diverse channel conditions while sustaining low latency in decision making.**

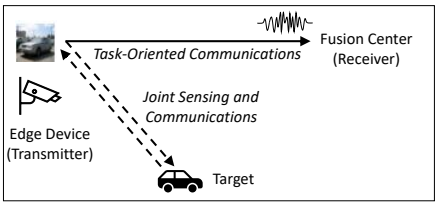
***Index Terms*—Integrated sensing and communications, task-oriented communications, deep learning, multi-modal data processing, dynamic spectrum access.**

I. Introduction

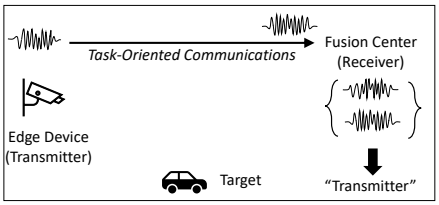
The increasing demand for wireless communication services has led to a spectrum scarcity challenge, highlighting the need for efficient spectrum management strategies. Dynamic Spectrum Access (DSA) has emerged as a promising solution to optimize spectrum utilization by allowing users to opportunistically access underutilized frequency bands. However, efficient implementation of DSA requires robust spectrum sensing techniques to identify and avoid interference with incumbent users. Spectrum sensing relies on collecting and analyzing wireless signals, which may be limited in discriminating between different types of transmitters and the environmental context in which they operate. Recently, image data has found ways to assist wireless tasks such as channel estimation and beam tracking [1]. The integration of diverse data modalities, such as image and wireless (spectrum) data, presents a novel and powerful approach to enhance overall sensing capabilities to support DSA and other spectrum tasks.

Recognizing the potential of *multi-modal data* for comprehensive situational awareness, we introduce a deep learning (DL) approach that synergistically leverages both image and target sensing data for the DSA purposes, as illustrated in Fig. 1. The proposed approach centers around an edge device equipped with a camera that collects image samples for object recognition, particularly focusing on identifying potential transmitters, such as those positioned at vehicles. One challenge is how to deliver multi-modal image and target sensing data efficiently from the edge devices to the receiver that needs to perform the task of transmitter identification. Acknowledging the communication and computational constraints and trust concerns associated with on-device computation, we consider a collaborative framework selectively transmitting processed information of much lower dimension from the edge device to a trusted receiver (a fusion center where the decision of transmitter identification is made). We employ the concept of *task-oriented communications* (TOC) or *goal-oriented communications* that optimizes the use of limited resources by tailoring the transmission to the specific task or goal of the receiver, avoiding unnecessary data transmission.

The primary goal of conventional communications has been the reliable delivery of information over a channel. *Autoencoder communications* [2], [3], involving the joint training of an encoder at the transmitter and a decoder at the receiver (both implemented as DNNs), can effectively adapt to channels with DL. Recently, *semantic communications* has emerged to preserve semantics (meaning) of delivered information [4]–[8], e.g., by ensuring that the received information leads to the correct outcome of a machine learning task [9], [10]. Semantics can be represented by the importance of the task to be completed, leading to the concept of TOC, where the objective is to complete a task at a receiver by using the



(a) Step 1: Edge device takes an image by its camera and sends it to the fusion center via TOC while this signal is also used for target sensing.



(b) Step 2: Edge device collects the signal reflected off the target and transmits it to the fusion center via TOC.

Fig. 1: Multi-modal data collection and delivery.

data samples originally available at the transmitter [11]–[16]. Instead of reconstructing all data samples, TOC delivers only a reduced amount of information to the receiver for completing the underlying task reliably and with low latency. TOC can serve multiple users each with different data modalities [17].

The resource-efficient approach of TOC is particularly advantageous in DSA scenarios, where quick and accurate decisions are imperative, and communication and computational resources are often constrained. For that purpose, we are employing an *encoder* at the transmitter to perform joint source coding, channel coding, and modulation, efficiently transmitting essential information of reduced dimension specifically tailored for object classification. This way, the receiver can reliably complete its task (i.e., make image classification decision) without reconstructing data samples.

In the meantime, the transmitted signals may interact with the environment and return to the transmitter, allowing for the collection of target sensing data. Traditionally, target sensing systems often resort to the transmission of dedicated probing signals to gather information about the spectral environment. This practice, while effective, comes at the cost of additional bandwidth consumption and potential interference with existing communication channels. Moreover, the deployment of extra probing signals may introduce delays and increase the complexity of the overall system.

Delivery of data such as images from the edge devices to the fusion center serves the dual purpose of target sensing. Recognizing these challenges, we leverage *integrated (joint) sensing and communications* (ISAC), where sensing (data collection or perception) and communication functionalities are unified [18]–[21]. Inclusion of radar-like sensing capabilities in communication networks is desired in DSA systems to gather spectrum information from the environment through sensors and communicate this information efficiently to a receiver without using additional probe signals. ISAC can be set up in form of autoencoder communications [22], [23] and can be integrated with semantic and TOC [24].

Fig. 1 shows the two steps of multi-modal data collection and delivery. In step 1, edge device takes an image and sends it to the fusion center via TOC. By combining target sensing and communication tasks in a unified learning framework, our system leverages the returned signals from the environment, originally transmitted for object recognition, as a valuable source of sensing data. This eliminates the necessity for additional probing signals, thereby conserving bandwidth resources and reducing the latency associated with target sensing.

In step 2, the collected target sensing data undergoes a second round of *encoding* at the transmitter, potentially condensing the information into a reduced-dimensional format for communication back to the fusion center through TOC. On the receiver side, a *decoder* plays a pivotal role in identifying active transmitters by fusing data received through both image transmission and target sensing. Importantly, the proposed system jointly trains the two encoders at the transmitter and the decoder at the receiver corresponding to three deep neural networks (DNNs). This approach enables a seamless integration of multi-modal data for image classification and wireless signal detection to identify potential transmitters for DSA. Note that this approach identifies the presence of transmitters even when they are passive (i.e., they are not transmitting) or their transmissions do not reach the spectrum sensors (e.g., because they are outside the transmission range or employ directional transmissions so that they cannot be captured by spectrum sensors) although they may reach and interfere with the receiver to be protected (such as the primary user receiver). In this context, the proposed approach aims to assist conventional spectrum sensing techniques that are employed to identify active transmissions.

To validate the efficacy of the proposed approach, we leverage the CIFAR-10 dataset as our source of images, creating a realistic representation of objects, including potential transmitters. Furthermore, we simulate communication and sensing in diverse channel conditions by incorporating Additive White Gaussian Noise (AWGN) and Rayleigh channel models. The results showcase high accuracy in transmitter identification across diverse channel conditions, while maintaining low latency for practical applicability in real-world DSA scenarios such as in the Citizens Broadband Radio Service (CBRS) band, where prompt and accurate decision-making is crucial for efficient spectrum utilization.

The remainder of the paper is organized as follows. Sec. II presents the joint sensing and TOC approach. Sec. III describes the encoder-decoder formulations. Sec. IV presents performance evaluation. Sec. V concludes the paper.

II. Joint Sensing and Task-Oriented Communications

The system model for joint sensing and TOC is shown in Fig. 2. Image data, initially transmitted for object recognition, is utilized for both image classification and target sensing. The following steps are pursued:

1) The transmitter (an edge device) captures images.

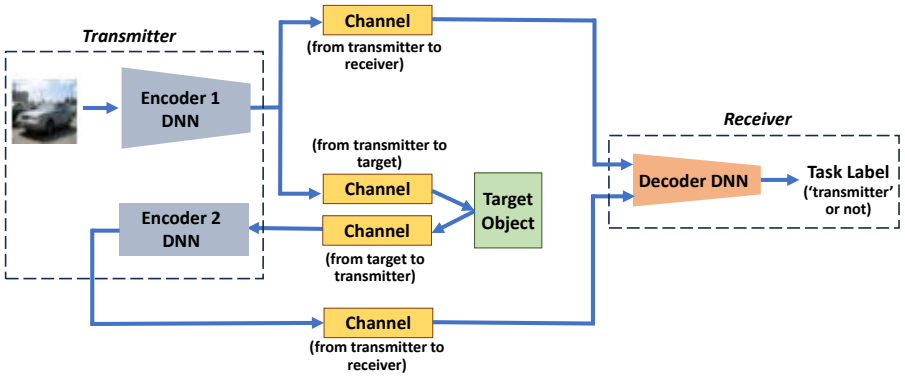


Fig. 2: System model for joint sensing and TOC.

2) The transmitter performs TOC by processing input data samples through an encoder to deliver the reduced amount of necessary information from the image samples to the receiver (fusion center).
3) This transmission is used for target sensing by letting the transmitted signals interact with the environment and return to the transmitter in an ISAC framework.
4) The transmitter performs another round of TOC by processing the reflected signals through another encoder to deliver the reduced amount of necessary information from the target sensing samples to the receiver.
5) The receiver collects signals from both rounds of TOC and process the fused data through a decoder to identify potential transmitters.

Image data at the transmitter. As the starting point, we consider an edge device that is equipped with a camera to take images. We use the CIFAR-10 dataset as a widely used collection of images designed for object recognition tasks in computer vision. The dataset consists of 60,000 32×32 color images. Each image sample has the dimension of $32 \times 32 \times 3 = 3072$ and belongs to one of ten distinct classes. These classes include common objects such as airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks. Each class contains 6,000 images. The dataset is evenly split into a training set of 50,000 images and a test set of 10,000 images.

To tailor the CIFAR-10 dataset for our specific task of distinguishing between potential transmitters (vehicles) and environmental objects (animals), we perform a binary classification by grouping the original ten classes into two labels: "Vehicles" and "Animals." Specifically, we merge the classes "airplanes", "automobiles", "ships", and "trucks" into the "Vehicles" label, encompassing transportation-related man-made objects. Conversely, the classes "birds," "cats," "deer," "dogs," "frogs," "horses," are grouped into the "Animals" label, representing a diverse set of natural objects within the environment. This binary classification allows us to focus on the distinction between potential transmitters associated with vehicles and other objects present in the surroundings. The DL model trained on this modified dataset then aims to efficiently differentiate between these two overarching categories, contributing to the identification of transmitters in DSA scenarios.

First round of TOC. The edge device that we refer to as the transmitter has the image data samples but they need to be classified at the receiver. We consider TOC that serves as a resource-efficient alternative to the conventional approach of delivering and reconstructing original data samples at the receiver. The data collected at the transmitter is encoded and selectively transmitted to the receiver. Rather than transmitting the data samples, TOC involves encoding information in a manner that aligns with the specific goal of the receiver (namely image classification).

An *encoder* at the transmitter performs joint source coding, channel coding, and modulation to efficiently represent the essential features needed for object classification. This reduced-dimensional information, tailored to the receiver's task of image classification, is transmitted over the channel. This approach significantly minimizes the amount of data that needs to be transmitted while retaining the critical information required for decision-making. At the receiver, a *decoder* interprets the received encoded information, extracting the features necessary for image classification. The decoder effectively reconstructs only the information relevant to the task at hand, bypassing the need for reconstructing the entire set of data samples. This targeted communication strategy not only conserves bandwidth but also divides computational load between the transmitter and the receiver, contributing to lower latency and more efficient decision-making. Finally, the transmitter as an edge device cannot be necessarily trusted to complete the task on its own using its encoder alone.

ISAC. This transmission can be also used for the dual purpose of target sensing in the framework of ISAC. The transmitted signals interact with objects in the environment, including potential transmitters. When these signals encounter objects, reflections occur. These reflections carry information about the objects and their surroundings. The reflected signals return to the transmitter. This return path involves the signals bouncing off objects, potentially modifying their characteristics based on the properties of the objects that they encounter. Depending on whether the signal is reflected from a vehicle or animal, reflected signals are received at the transmitter with different signal-to-noise ratios (SNRs), in particular smaller sensing SNR for animals compared to vehicles due to more absorption.

The sensing SNR incorporates the effects of channels to and from the target and the effects of reflections from the target.

Second round of TOC. The transmitter sends these reflected target sensing signals to the receiver by another round of TOC. A second *encoder* is trained at the transmitter to process the target sensing samples before transmitting them to the receiver.

Transmitter identification at the receiver. The receiver processes multi-modal data from two rounds of TOC, the first one for transmission of image data and the second one for transmission of target sensing data. For that purpose, the receiver employs a *decoder* to classify the signals (collected and combined over two rounds of TOC) into two labels, namely potential transmitter or not. This decoder at the receiver is jointly trained with the two encoders at the transmitter to minimize the loss between the true and predicted labels.

Channels. We consider two types of channel models for both communication from the transmitter to receiver as well as for target sensing. First, we consider the AWGN channel model, in which the received signal at the receiver consists of the transmitted signal along with white Gaussian noise. Second, we consider the Rayleigh fading channel, where the transmitted signal undergoes Rayleigh fading, and white Gaussian noise is added to the received signal. Channel distributions do not change during training and testing phases, even as individual outcomes may differ and are not known beforehand.

III. Encoder-Decoder Formulations

In this section, we establish how the encoders and the decoders are connected to form the input-output relationships for the underlying DNNs. Suppose that the transmitter (T) has $\boldsymbol{x}$ as the input (image) data samples. Let $\boldsymbol{h}_{ij}^{(k)}$ denote the channel from node i to node j and $\boldsymbol{n}_j^{(k)}$ denote the noise at node j at the k-th round of TOC. We assume that the channel and noise distributions are the same over the two rounds of TOC, while their realizations will change. T encodes $\boldsymbol{x}$ as $E_1(\boldsymbol{x})$ and transmits it over the channel to the receiver (R). The signal that is received by R is given by

$$\boldsymbol{y}_R^{(1)} = \boldsymbol{h}_{TR}^{(1)} E_1(\boldsymbol{x}) + \boldsymbol{n}_R^{(1)} \tag{1}$$

at the first round of TOC. In the meantime, the transmitted signal $E_1(\boldsymbol{x})$ is reflected off the object (O), namely the target, and returns back to T as

$$\boldsymbol{y}_T^{(1)} = \boldsymbol{h}_{TOT}^{(1)} E_1(\boldsymbol{x}) + \boldsymbol{n}_T^{(1)}, \tag{2}$$

where $\boldsymbol{h}_{TOT}^{(1)}$ is the combined channel from T to O and from O to T. Then T encodes $\boldsymbol{y}_T^{(1)}$ as $E_2\big(\boldsymbol{y}_T^{(1)}\big)$ and transmits it to R. At the second round of TOC, the signal received by R is

$$\begin{aligned} \boldsymbol{y}_R^{(2)} &= \boldsymbol{h}_{TR}^{(2)} E_2\big(\boldsymbol{y}_T^{(1)}\big) + \boldsymbol{n}_R^{(2)} && (3)\\ &= \boldsymbol{h}_{TR}^{(2)} E_2\big(\boldsymbol{h}_{TOT}^{(1)} E_1(\boldsymbol{x}) + \boldsymbol{n}_T^{(1)}\big) + \boldsymbol{n}_R^{(2)}. && (4) \end{aligned}$$

The signals received at the two rounds of TOC are combined at R as $\boldsymbol{y}_R = \big\{\boldsymbol{y}_R^{(1)}, \boldsymbol{y}_R^{(2)}\big\}$ and processed through the decoder D to return the predicted label $\hat{l}$ that is given by

$$\begin{aligned} \hat{l} = D\big(\boldsymbol{y}_R\big) &= D\big(\big\{\boldsymbol{y}_R^{(1)}, \boldsymbol{y}_R^{(2)}\big\}\big) && (5)\\ &= D\big(\big\{\boldsymbol{h}_{TR}^{(1)} E_1(\boldsymbol{x}) + \boldsymbol{n}_R^{(1)}, \\ &\qquad \boldsymbol{h}_{TR}^{(2)} E_2\big(\boldsymbol{h}_{TOT}^{(1)} E_1(\boldsymbol{x}) + \boldsymbol{n}_T^{(1)}\big) + \boldsymbol{n}_R^{(2)}\big\}\big). && (6) \end{aligned}$$

A benchmark for comparison is the case of target sensing only when D takes $\boldsymbol{y}_R = \big\{\boldsymbol{y}_R^{(2)}\big\}$ as the input and returns the predicted label

$$\hat{l} = D\big(\boldsymbol{h}_{TR}^{(2)} E_2\big(\boldsymbol{h}_{TOT}^{(1)} E_1(\boldsymbol{x}) + \boldsymbol{n}_T^{(1)}\big) + \boldsymbol{n}_R^{(2)}\big). \tag{7}$$

The overall DNN structures used in (6) and (7) combine the two encoders E_1 and E_2, and the decoder D along with the channel effects. E_1, E_2, and D are jointly trained by accounting for all the underlying channel and noise effects $\boldsymbol{h}_{TR}^{(1)}, \boldsymbol{h}_{TR}^{(2)}, \boldsymbol{h}_{TOT}^{(1)}, \boldsymbol{n}_R^{(1)}, \boldsymbol{n}_T^{(1)}$, and $\boldsymbol{n}_R^{(2)}$.

The DNN architectures used for the two encoders and the decoder are given in Table I, where $n_{c,1}$ and $n_{c,2}$ are the output sizes of encoders 1 and 2, respectively. The input-output relationships and interactions of DNNs are shown in Fig. 3. We consider a convolutional neural network (CNN) architecture that is separated between the two encoders and the decoder. Simpler architectures such as the feedforward neural networks (FNNs) are known to fall short of capturing the complexity of CIFAR-10 dataset. Hyperparameters are selected by gradually increasing the number of layers and layer sizes to determine the configuration with the best performance.

TABLE I: Properties of DNN architectures for joint sensing and TOC.

Network	Layer	Properties
Encoder 1	Input	size: 32×32×3
	Conv2D	filter size: 8, kernel size: (3,3) activation: ReLU
	Conv2D	filter size: 4, kernel size: (3,3) activation: ReLU
	MaxPooling2D	pool size: (2,2)
	Dropout	dropout rate: 0.1
	Conv2D	filter size: 4, kernel size: (3,3) activation: ReLU
	MaxPooling2D	pool size: (2,2)
	Dropout	dropout rate: 0.1
	Flatten	–
	Dense	size: 128, activation: ReLU
	Dense	size: $n_{c,1}$, activation: Linear
Encoder 2	Input	size: $n_{c,1}$
	Dense	size: $n_{c,1}$, activation: ReLU
	Dense	size: $\frac{1}{2}(n_{c,1}+n_{c,2})$, activation: ReLU
	Dense	size: $n_{c,2}$, activation: Linear
Decoder	Input	size: $n_{c,1}+n_{c,2}$
	Dense	size: $n_{c,1}+n_{c,2}$, activation: ReLU
	Dense	size: $\frac{1}{2}(n_{c,1}+n_{c,2})$, activation: ReLU
	Dense	size: 2, activation: Softmax

IV. Performance Evaluation

We obtain numerical results through Python, and the DL model is trained using Keras with the TensorFlow backend. Categorical cross-entropy loss is minimized as the loss between true labels l and predicted labels $\hat{l}$ (given by (6) or (7)). The optimizer employed is Adam. The batch size is 64. The number of epochs for training is 5. The resulting DNN structure is trained with 50,000 data samples and tested with 10,000 data samples. We consider the following default values of the system parameters. The SNR for communication

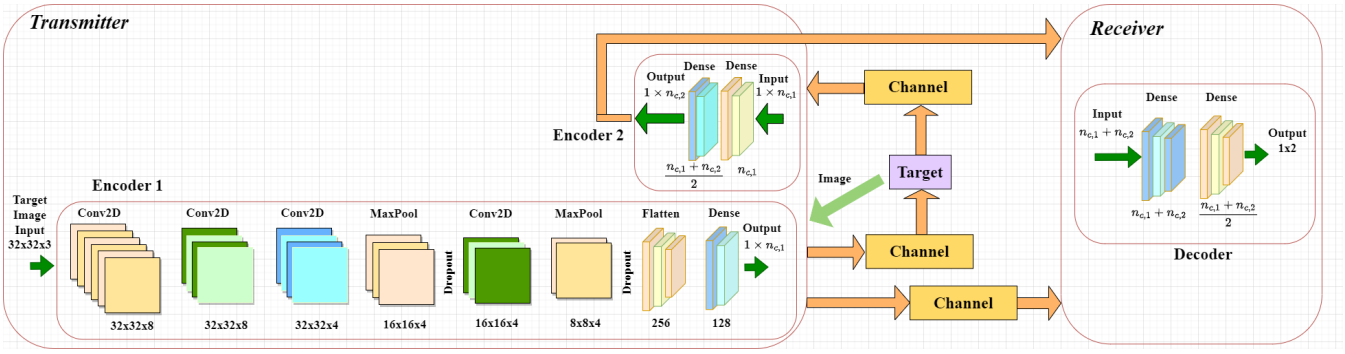
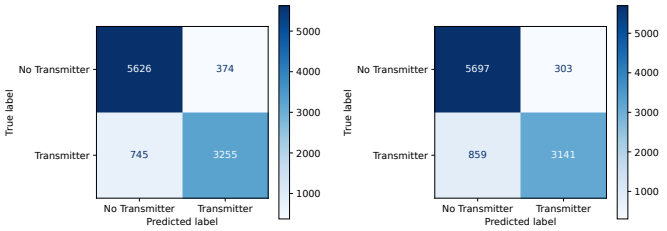


Fig. 3: Encoder and decoder DNNs for joint sensing and TOC.



(a) AWGN channel. (b) Rayleigh channel.

Fig. 4: Confusion matrix for default system values.

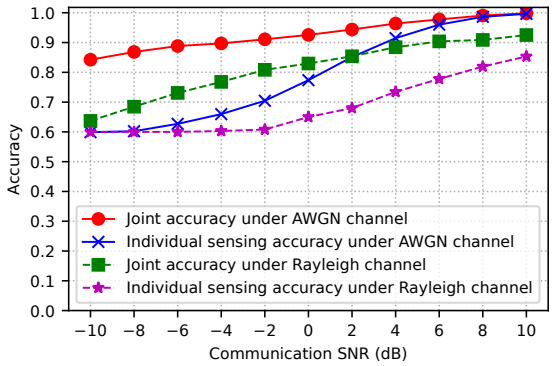


Fig. 5: Accuracy vs. SNR_c (dB), when SNR_s is 6dB lower.

channel, SNR_c, from the transmitter to the receiver is 3dB. The SNR for sensing, SNR_s, combines the effect of channels from the transmitter to the target and back from the target to the transmitter and the effect of reflections from the targets depending on whether they are vehicles or animals. We set the maximum SNR_s for vehicles as -3dB and the maximum SNR_s for animals 6dB lower. The size of encoder output is 20, which corresponds to the concatenation of the in-phase and quadrature components of complex symbols. Since each image sample has the dimension of 3072, this corresponds to compression rate of 0.65%, indicating highly efficient use of communication resources. For computational complexity, there are $38,704$ parameters to train in the encoder and decoder DNNs. The time spent for transmission is $(n_{c,1}+n_{c,2})\frac{\tau}{2}$, where τ is the time to transmit one complex symbol.

Under default system parameters, the accuracy of joint sensing and TOC is 0.97 and 0.88 under the AWGN and Rayleigh channels, respectively. Fig. 4 shows the confusion matrix under the AWGN and Rayleigh channels for 10,000 test samples. The decoder makes fewer errors in identifying the absence of a transmitter rather than the presence of a transmitter (i.e., false alarm probability is lower than misdetection probability).

Next, we vary the system parameters one by one by fixing the rest to default values. Fig. 5 shows the accuracy as a function of SNR_c (dB), when SNR_s is set 6dB lower than SNR_c. The joint accuracy refers to the accuracy achieved by joint sensing and TOC, whereas the individual sensing accuracy refers to the accuracy achieved by providing signals reflected off the transmitter as the only input to the decoder. Results show that the joint accuracy is higher than the individual sensing accuracy for AWGN and Rayleigh channels with different SNR_c. Overall, the accuracy is higher under the AWGN channel compared to the Rayleigh channel. As SNR_c increases, the accuracy increases in all cases. The individual sensing accuracy catches up with the joint sensing accuracy for high SNR_c such as 10dB under the AWGN channel, whereas the gap closes slowly with SNR_c under the Rayleigh channel.

Fig. 6 shows the accuracy as a function of SNR_s (dB), when SNR_c is fixed to 3dB. The effect of SNR_s is observed more on the individual sensing accuracy compared to the joint accuracy. The gap between the joint sensing accuracy and individual sensing accuracy closes as SNR_s increases (more under the AWGN channel than the Rayleigh channel). Overall, the accuracy is also higher under the AWGN channel compared to the Rayleigh channel.

Fig. 7 shows the accuracy as a function of transmitter output size for both encoders (where $n_{c,1} = n_{c,2} = n_c$). The joint accuracy remains high even when the n_c is small (which corresponds to high compression of transmitted signals and efficient use of channel resources). In all cases, the accuracy improves with n_v and starts saturating without n_c growing significantly. Again, the accuracy is higher under the AWGN channel than the Rayleigh channel. While the individual sensing accuracy quickly approaches the joint sensing accuracy as SNR_s increases under the AWGN channel, the accuracy gap remains high under the Rayleigh channel.

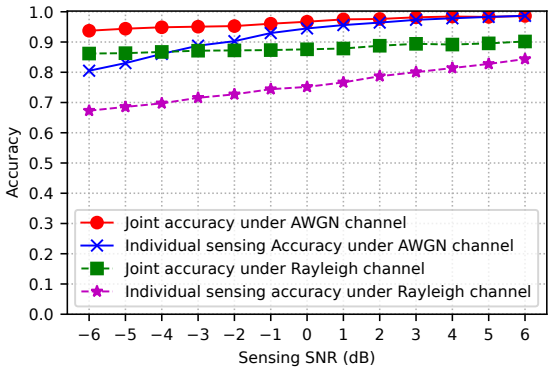


Fig. 6: Accuracy vs. SNR_s (dB), when SNR_c is 3dB.

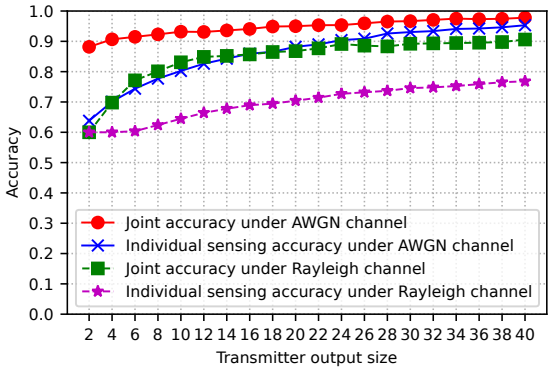


Fig. 7: Accuracy vs. transmitter output size.

V. Conclusion

We have presented a novel approach to transmitter identification by harnessing the power of multi-modal image and target sensing data integration. Focusing on an edge device equipped with a camera to capture images, the system addresses computational constraints and trust issues associated with on-device processing. The edge device selectively communicates processed information to a trusted receiver acting as a fusion center. TOC facilitates efficient transmission of reduced-dimensional information for object classification. Concurrently, the system leverages reflections of transmitted signals to collect target sensing data via ISAC. The integration of two rounds of encoding at the transmitter and joint training of two encoders at the transmitter and a decoder at the receiver enables seamless fusion of multi-modal data. This approach achieves high accuracy in transmitter identification for both AWGN and Rayleigh channels and sustains low latency by the elimination of additional probing signals and the resource-efficient transmission of reduced-dimensional information.

References

[1] A. Alkhateeb, G. Charan, T. Osman, A. Hredzak, J. Morais, U. Demirhan, and N. Srinivas, "Deepsense 6G: A large-scale real-world multi-modal sensing and communication dataset," *IEEE Communications Magazine*, vol. 61, no. 9, pp. 122–128, 2023.

[2] T. J. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[3] T. Erpek, T. J. O'Shea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep learning for wireless communications," *Development and Analysis of Deep Learning Architectures*, pp. 223–266, 2020.

[4] B. Güler, A. Yener, and A. Swami, "The semantic communication game," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 787–802, 2018.

[5] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2022.

[6] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next generation semantic communication networks," *arXiv preprint arXiv:2211.14343*, 2022.

[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] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, T. Soleymani, B. S. Soret, and H. Johansson, "Semantic communications in networked systems," *IEEE Network*, vol. 36, no. 4, pp. 233–240, 2022.

[9] Y. E. Sagduyu, T. Erpek, S. Ulukus, and A. Yener, "Is semantic communications secure? a tale of multi-domain adversarial attacks," *arXiv preprint arXiv:2212.10438*, 2022.

[10] Y. E. Sagduyu, T. Erpek, A. Yener, and S. Ulukus, "Will 6G be semantic communications? opportunities and challenges from task oriented and secure communications to integrated sensing," *arXiv preprint arXiv:2401.01531*, 2024.

[11] H. Zhang, S. Shao, M. Tao, X. Bi, and K. B. Letaief, "Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 170–185, 2022.

[12] Y. E. Sagduyu, S. Ulukus, and A. Yener, "Task-oriented communications for nextG: End-to-end deep learning and AI security aspects," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 52–60, 2023.

[13] ——, "Age of information in deep learning-driven task-oriented communications," in *IEEE INFOCOM Age of Information Workshop*, 2023.

[14] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.

[15] X. Kang, B. Song, J. Guo, Z. Qin, and F. R. Yu, "Task-oriented image transmission for scene classification in unmanned aerial systems," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5181–5192, 2022.

[16] Y. Shi, Y. Zhou, D. Wen, Y. Wu, C. Jiang, and K. B. Letaief, "Task-Oriented Communications for 6G: Vision, Principles, and Technologies," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 78–85, 2023.

[17] 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.

[18] U. Demirhan and A. Alkhateeb, "Integrated sensing and communication for 6G: Ten key machine learning roles," *arXiv preprint arXiv:2208.02157*, 2022.

[19] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1728–1767, 2022.

[20] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under gaussian channels," *IEEE Transactions on Information Theory*, vol. 69, no. 9, pp. 5723–5751, 2023.

[21] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. P. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, pp. 1295–1315, 2021.

[22] J. M. Mateos-Ramos, J. Song, Y. Wu, C. Häger, M. F. Keskin, V. Yajnanarayana, and H. Wymeersch, "End-to-end learning for integrated sensing and communication," in *IEEE International Conference on Communications*, 2022.

[23] C. Muth and L. Schmalen, "Autoencoder-based joint communication and sensing of multiple targets," in *International ITG Workshop on Smart Antennas and Conference on Systems, Communications, and Coding*, 2023.

[24] Y. E. Sagduyu, T. Erpek, A. Yener, and S. Ulukus, "Joint sensing and semantic communications with multi-task deep learning," *arXiv preprint arXiv:2311.05017*, 2023.