This is a peer-reviewed, author's accepted manuscript of the following conference output: (Eds.) (2024).
Semantic communication based complexity scalable image transmission system for resource constrained devices.
Paper presented at 2024 IEEE CTSoc Gaming, Entertainment and Media (GEM) Conference, Turin, Italy. Advance online publication.

Semantic Communication Based Complexity Scalable Image Transmission System for Resource Constrained Devices

Prabhath Samarathunga
Department of Computer and Information Sciences
University of Strathclyde
Glasgow, UK
prabhath.samarathunga@strath.ac.uk

Yasith Ganearachchi
Department of Computer and Information Sciences
University of Strathclyde
Glasgow, UK
yasith.ganearachchi@strath.ac.uk

Thanuj Fernando
Department of Computer and Information Sciences
University of Strathclyde
Glasgow, UK
thanuj.fernando.2023@uni.strath.ac.uk

Lahiru Thanippulige
Department of Computer and Information Sciences
University of Strathclyde
Glasgow, UK
lahiru.thanippulige@strath.ac.uk

Anil Fernando
Department of Computer and Information Sciences
University of Strathclyde
Glasgow, UK
anil.fernando@strath.ac.uk

***Abstract*—Multimedia traffic is expanding at an astonishing rate and is finding its way into many new applications, including wireless sensor networks and devices from the Internet of Things. However, the bandwidth and energy requirements associated with them are becoming increasingly prohibitive, challenging the sustainability of many communication networks. Semantic communications offer a novel approach to overcome bandwidth limitations, but the complexity of encoders and decoders limits their application in resources-constrained devices. We propose a semantic communication-based complexity-scalable image transmission system that uses asymmetric autoencoders to shift the complexity of the system to the encoder or decoder without compromising on human and machine perceived quality. Using a test data set, we demonstrate that the complexity can be shifted between the two without affecting the overall performance of the semantic communication system. This concept will be significant in the implementation of wireless sensor networks based on semantic communication and Internet of Things applications, while also providing a novel tool to further improve conventional encoder simplification approaches, such as distributed coding. Furthermore, the realization of energy efficient sensors by utilizing semantic communications and complexity scalability will have a direct impact on achieving sustainable development goals related to energy usage and management in media communications.**

***Index Terms*—Asymmetric Autoencoders, Image Communications, IoT, Semantic Communications, Wireless Sensor Networks**

I. Introduction

Multimedia communication has become an essential part of many applications ranging from entertainment, health, transportation, agriculture, aquaculture, remote sensing, security, and smart cities. However, the increasingly complex nature of media and communication systems has led to an enormous increase in its bandwidth and energy demands, becoming a major concern for the sustainability of the entire ecosystem. The widespread adaptation of sustainability concepts across such systems and a significant overall reduction in the negative impact of resource utilization and the environment are key concerns as we move into the future.

Sustainable information communication technology, or *Green ICT* [1], [2], covers the creation, use, management, and disposal of devices and systems that minimize environmental impact, as shown in Fig. 1. This has led to more investments towards smart cities that contain smart transportation, open data, smart buildings, smart manufacturing, and smart citizens governed by smart governments. This growing dependence on information communication technology in our daily lives must be managed to reduce the dependency on environmental resources, while minimizing the pollution caused by their improper use.

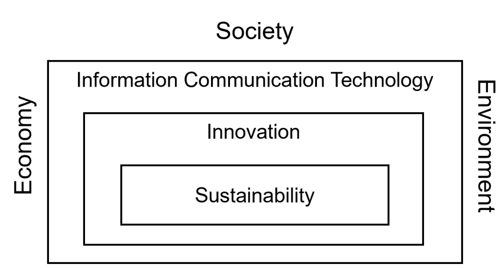


Fig. 1. The Concept of Green ICT

Battery-operated meters, thermostats in smart home concepts, smart assets, pet tracking devices, smart crop sensors, connected cellular smart meters, and even sensor networks of electric mobile vehicles such as cars or scooters fall into the category of low-power devices under IoT, as do common appliances such as televisions, refrigerators, mobile communication devices, and even personal computers. The application of reducing power consumption is often utilized to improve energy usage. This will be due to the inability to power the IoT gadgets, as well as the exclusive usage of the battery. Selecting an appropriate IoT may lead to ensuring the stability and high reliability of users and the future of a sustainable environment [3].

Novel approaches for modeling resource, material, manufacturing, and power management are necessary to implement this vision. The systemic nature of these systems means that they are ideally placed to undergo a complete reformation in their practices. A key requirement for this transformation is the minimization of bandwidth and energy footprint of wireless sensor networks (WSN) and the Internet of Things (IoT) as 5G and 6G systems under ever-increasing traffic load with such surging multimedia traffic. These services and networks are also supposed to support smart and adaptive operation using complex control frameworks, even increasing resource consumption.

A novel approach to the communication problem in such networks can make use of semantic communications. First described in [4], it suggests that, given the existence of a shared context between a transmitter and a receiver, the semantic or meaning of a message alone is sufficient to achieve the desired effect of communication and even reconstruct the original information. It is expected to reduce bandwidth and complexity while increasing range and enabling longer operational cycles in battery-powered devices for WSN and IoT. However, these semantic communication systems suffer from requiring significant computing resources to train and run compared to conventional communication systems, limiting their application in scenarios where either the encoder or decoder is resource-constrained, mainly in terms of computational power and bandwidth.

We propose a system of image transmission based on asymmetric autoencoder semantic communication scalable by complexity. The optimization of the Deep Neural Network used in the proposed system to achieve the performance balance between processing complexity and bandwidth utilization for image transmission applications. The major impact of the proposed research will be the application of the proposed concept of semantic communication to collect and analyze sensor outputs in a semantic context (as opposed to a physical context, which is done now), enabling machines to identify the context of the environment they sense and perform machine-based decision making, which can greatly increase the value of different types of state-of-the-art WSN and IoT-based solutions. The proposed work may well be a major breakthrough that enables the capabilities envisaged for 6G in future WSN and IoT.

II. Related Work

The general communication system for electronic transmission of information was first introduced along with 11 fundamental theorems for understanding it [5] . The last of these theorems states that for a channel with capacity C and a discrete information source with entropy per second H there is no encoding method that gives an equivocation less than $H - C$. This effectively imposes a maximum rate of error-free transmission on a channel subject to electronic noise as shown in (1), where C is the channel capacity, B is the channel bandwidth, and $\frac{S}{N}$ is the signal-power-to-noise-power ratio of the signal.

$$C = B \log_2 \left(1 + \frac{S}{N}\right) \tag{1}$$

Although digital communication has evolved exponentially since [5], this bound still governs the maximum data rate that can be reached on any channel.

Semantic communication has recently reemerged as a way to circumvent this limit. It was first discussed in suggesting the existence of three problems in communication systems: *technical* (“How accurately can the symbols of communication be transmitted?”), *semantic* (“How precisely do the transmitted symbols convey the desired meaning?”), and *effectiveness* (“How effectively does the received meaning affect conduct in the desired way?”) [4]. Conventionally, resolution of *effectiveness* and *semantic* problems can only be achieved if the *technical* problem is successfully resolved, that is, communication symbols are transmitted accurately. This is typically ensured by using source and channel coding, which improves the resilience of a signal against errors induced by channel noise.

Recent advances in artificial intelligence/machine learning that enable efficient and effective implementation of deep neural networks have made it possible to reliably extract the *semantic* of a message, which can be thought of as a highly compressed version of the original message. If a transmitter and a receiver share the context of how the semantics was extracted, the receiver can use it to reconstruct the original message with a high degree of precision simply by sharing the *semantic* between them [6], [7]. This concept has been successfully demonstrated for use in text transmission [8], image transmission [9] and video transmission [10] using a range of deep neural networks such as convolutional neural networks (CNN), generative adversarial networks (GAN) and autoencoders (AE). Currently attracting wide research interest, semantic communications have been identified as one of the key enablers for the next generation of mobile communication system (6G) with applications in the Metaverse [11], IoT [12] and even for infrastructure capabilities such as mobile edge computing (MEC) [13].

Only requiring a very small amount of data while providing significantly better resilience to channel noise compared to conventional encoding methods, as demonstrated in [14] by comparing the bits-per-pixel (bpp) requirement of a semantic communication based image transmission system and the Joint Photographic Experts Group (JPEG) codec, semantic communication based media transmission is ideal for use in low bandwidth applications such as WSN and IoT devices. However, a major restriction of its practical implementation in such applications is the computational complexity introduced by deep neural networks used for semantic extraction (by deriving a latent vector from the message) at the transmitter and reconstruction of the message (from the received latent vector) at the receiver, as demonstrated by the complexity of the encoders used by [8]–[10]. This prohibits its use in WSN applications, where the transmitters are usually resource constrained in terms of bandwidth, energy, and processing power, and in IoT applications, where both the transmitter and receiver can be resource constrained over the same attributes. This will directly affect the sustainability of any solution developed using semantic communication-based media transfer.

A similar challenge exists in conventional media encoding systems, and various techniques have been investigated to shift computational complexity from the transmitter or encoder side to the receiver or decoder side. They are more common in medical imaging, as discussed in [15] and [16], and in industrial and aerospace applications such as [17]. For media content, investigations have been carried out on distributed coding, as described in [18], where the complexity of the encoding could be moved from the transmitter to the receiver, reducing the cost of the transmitter, as demonstrated in [19].

However, since semantic communication-based media transmission systems are still in their infancy, no similar attempts have been reported in the literature. This is an initial attempt to use the features offered by deep neural networks, such as asymmetric autoencoders, to asymmetrically distribute the complexity of the end-to-end semantic communication system so that the end-to-end reference system can be reconfigured to have complex encoders and simple decoders or simple encoders and complex decoders, as shown in Fig. 2, according to the application scenario.

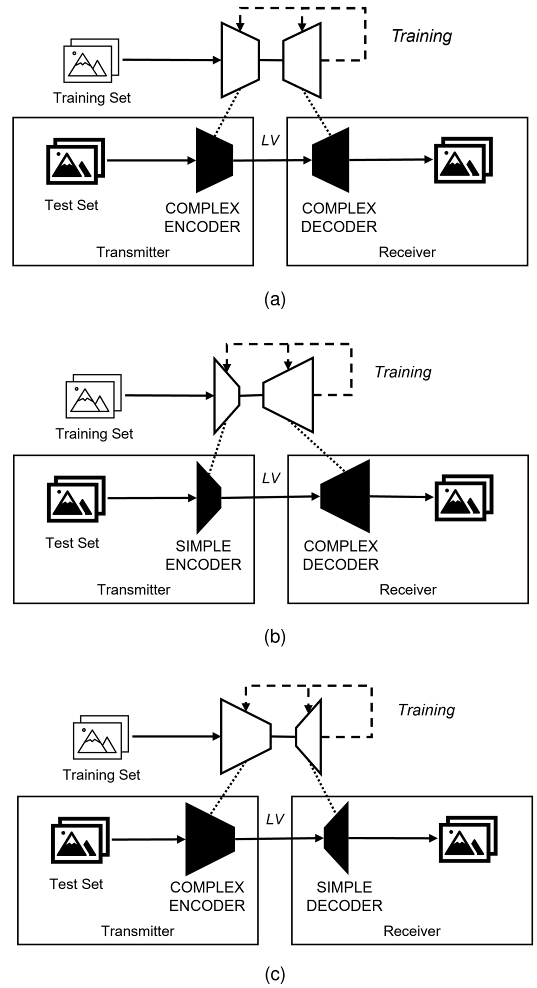


Fig. 2. Complexity Scalable Semantic Image Communication System Using Asymmetric Autoencoders. (a) Reference Model, (b) Simplified Encoder Model, (c) Simplified Decoder Model.

III. Proposed System

Semantic communication-based systems for media communication are usually implemented using symmetric deep neural networks, such as the undercomplete autoencoder shown in Fig. 3a to encode and decode images after being trained using a suitable data set, where the encoder layers are used at the transmitter and the decoder layers are used at the receiver. The latent vector created in the network bottleneck represents the *semantic* of the input image, and the hyperparameters of the encoder and decoder layers constitute the shared context between the transmitter and the receiver. We propose the use of an asymmetric autoencoder network, which can be configured as the simple encoder and complex decoder setup shown in Fig. 3b, or as the complex encoder and simple decoder setup shown in Fig. 3c. This would enable adapting the semantic communication system to resource-constrained transmitting devices and to receiving devices, as simple encoder and simple decoder configurations implemented with asymmetric autoencoders require significantly less computational complexity compared to the use of symmetric autoencoders.

The proposed system is tested using the Modified National Institute of Standards and Technology (MNIST) dataset [20], which is a large database of handwritten digits commonly used to train various image processing systems. The MNIST dataset comprises 28×28 pixel grayscale images of handwritten digits from 0 to 9. For the purpose of this study, a subset of the MNIST dataset was used, comprising 20,000 images for training and 10,000 images for testing. This selection of subsets was designed to evaluate the effectiveness of asymmetric autoencoders in a controlled environment.

The autoencoder network used for testing consists of multiple layers, each using a Leaky Rectified Linear Unit (LeakyReLU) with an alpha parameter set to 0.01 for the activation function, while the final layer employs a sigmoid activation function to ensure that the output values are normalized between 0 and 1, conducive to binary image reconstruction. The AE is trained using the Adam optimization algorithm with a learning rate of 0.001. The loss function used during the training process is the mean squared error (MSE). Training is carried out over 50 epochs in batches of 128 images each, which was determined to be sufficient for convergence without overfitting given the size and complexity of the MNIST dataset. The network bottleneck was set at 32 bytes, which is the latent vector representing the extracted *semantic* of the input images.

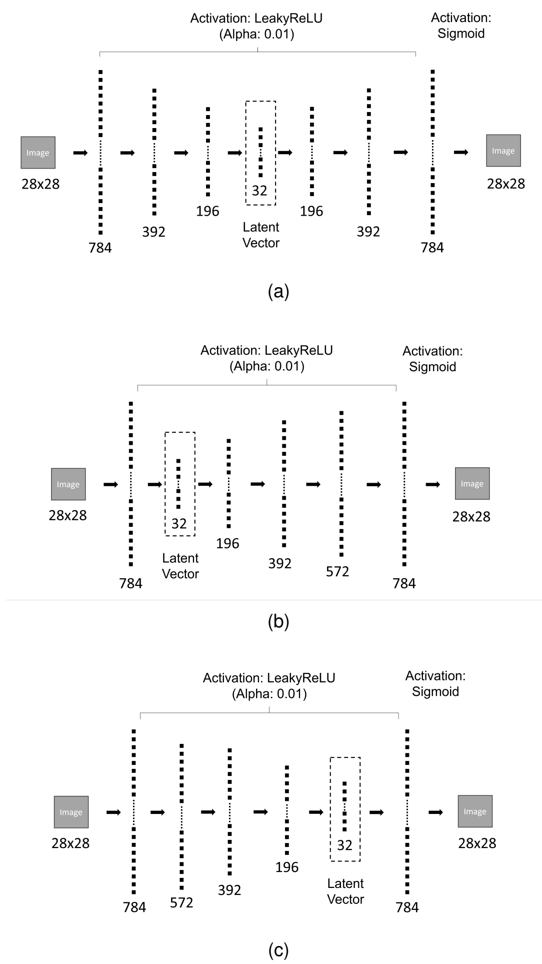


Fig. 3. Autoencoder Architectures Used for Testing the Complexity Scalable Semantic Encoder. (a) Reference system, (b) Simplified Encoder System, (c) Simplified Decoder System.

Evaluation of the practical utility of the reconstructed images is done using an independently trained convolutional neural network specifically designed for digit classification and previously validated using the MNIST dataset to serve as a benchmark for validation accuracy. After the autoencoder training process, the reconstructed test images were fed into the CNN model to determine the classification accuracy.

The architecture constitutes a sequentially configured convolutional neural network for the classification of the 28×28 pixel images from the MNIST dataset. It begins with a convolutional layer consisting of 32 filters of size 3×3, using the ReLU activation function, which is immediately followed by a max pooling layer with a 2×2 window to reduce dimensionality. A second convolutional layer with 64 filters, also followed by a 2×2 max pooling layer, continues the feature extraction process. Subsequently, the network transitions from convolutional operations to fully connected layers, facilitated by a flattening layer, converting the 2D feature maps into a 1D feature vector. This vector feeds into a dense layer of 128 neurons, followed by a second dense layer comprising 64 neurons, both employing the ReLU activation function for non-linear transformations. The final layer is a densely connected layer with 10 neurons, each corresponding to one of the ten possible digit classes in the MNIST dataset, utilizing a sigmoid activation function well suited for binary classification tasks, as it ensures that the output is between 0 and 1, which can be interpreted as the probability that the input is a particular digit. The classification network, with its two convolutional and max pooling layers followed by three densely connected layers, is designed to capture hierarchical patterns within the digit images, enabling effective learning and classification. It has a total of 232,650 trainable parameters and is compiled with the Adam optimizer and binary cross-entropy loss function, indicative of its applicability to multilabel classification scenarios. The training process is structured to run over 10 epochs with a batch size of 32.

To provide a comprehensive evaluation of the performance of the proposed system, we used a multifaceted approach using several image quality assessment metrics, namely the Peak Signal-to-Noise Ratio (PSNR), Root Mean Square Error (RMSE), Universal Quality Index (UQI), and Mean Squared Error (MSE).

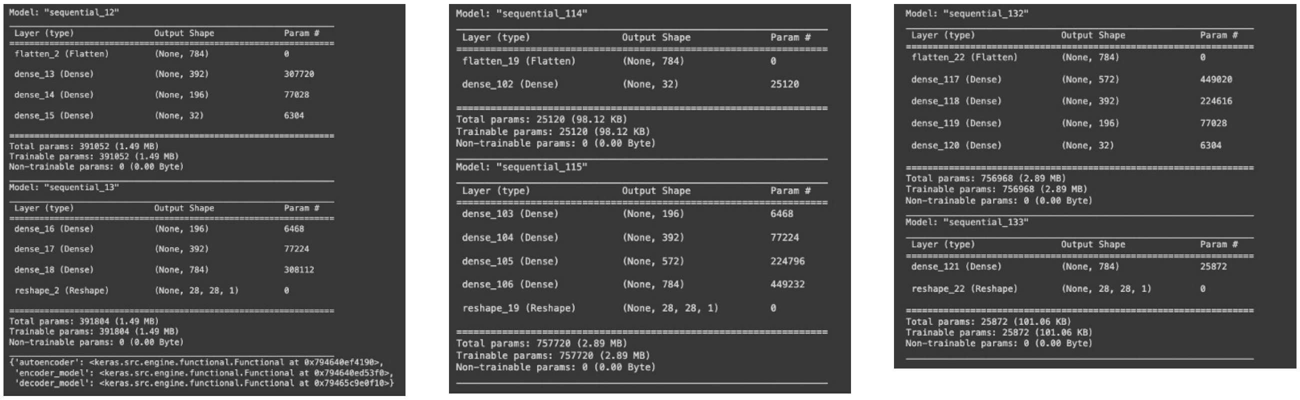
IV. Results and Discussion

To demonstrate the concept, we transmitted the same image data sets over the three autoencoder configurations shown in Fig. 3. The simple encoder (Fig. 3b) and the simple decoder (Fig. 3b) are considered such that the number of floating point operations between all codecs is as close as possible. Table I summarizes the number of floating point operations in the three cases.

TABLE I
Complexity of Encoder and Decoder Networks for the Three Configurations based on Number of Tunable Parameters (Floating Point Operations)

Configuration	Encoder	Decoder	Total
Reference	391,052	391,804	782,856
Simple Encoder	25,120	757,720	782,840
Simple Decoder	756,968	25,872	782,840

Fig. 4 shows the complexity of the three models and how it is distributed between the encoder and the decoder. The reference model, with a balanced distribution of parameters between the encoder and the decoder (Fig. 4a) has a complexity of $\mathcal{O}(391,052)$ at the encoder and $\mathcal{O}(391,804)$ at the decoder. The simple encoder-complex decoder configuration (Fig. 4b), shifts most of the complexity to the decoder and has a complexity of $\mathcal{O}(25,120)$ at the encoder and $\mathcal{O}(757,720)$ at the decoder. The complex encoder-simple decoder configuration



(a) (b) (c)

Fig. 4. Complexity Metrics of the Proposed Systems: (a) Reference (Symmetric) System (Model *sequential_12* corresponds to the encoder and model *sequential_13* corresponds to the receiver), (b) Simplified Encoder (Model *sequential_114* corresponds to the encoder and model *sequential_115* corresponds to the receiver), (c) Simplified Decoder (Model *sequential_132* corresponds to the encoder and model *sequential_133* corresponds to the receiver)

(Fig. 4c) shifts most of the complexity to the encoder and has a complexity of $\mathcal{O}(756, 968)$ at the encoder and $\mathcal{O}(25, 872)$ at the decoder. Fig. 5 shows the model loss performance of the three systems and demonstrates that the simple encoder and simple decoder configurations with asymmetric autoencoders converge in a manner similar to the reference configuration with a symmetric autoencoder.

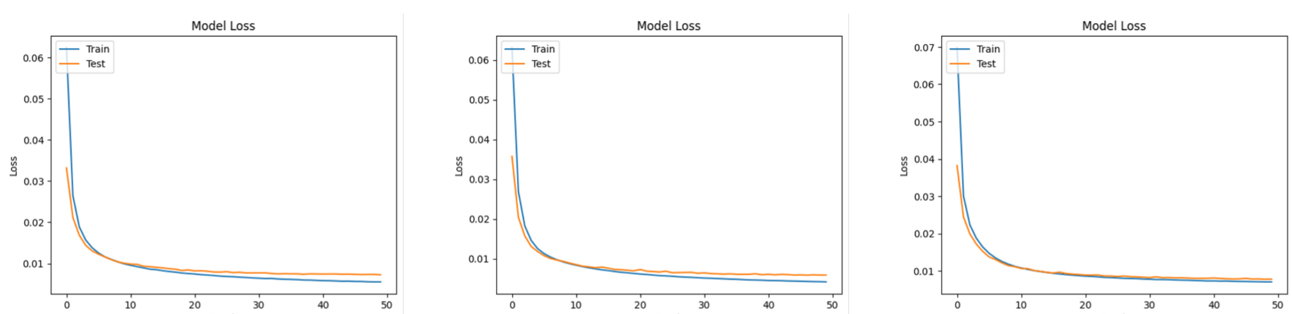
Table II summarizes the key quality parameters that compare transmitted and received images between different configurations. It should be noted that the comparatively low PSNR values are the result of the model being optimized for the identification of handwritten digits rather than for full reconstruction of the input image. The test accuracy of the three configurations is nearly equivalent, as are the four quality parameters. It was also observed that the results from the simple encoder configuration are better than those from the simple decoder or even the reference configurations. This is due to the hyperparameters of the decoder network having more influence in providing a more precise and accurate representation of the data, as discussed in [21].

Our findings suggest that the asymmetric autoencoder architectures used in simple encoder-complex decoder and complex encoder-simple decoder configurations can achieve high performance in both reconstruction quality and classification accuracy. The simple encoder-complex decoder configuration exhibits superior performance in most metrics, particularly in classification accuracy and PSNR, indicating its potential for applications where model efficiency and high-quality reconstruction are desired. These insights could pave the way for the development of more efficient deep neural network architectures that maintain high performance while being computationally less demanding, which is of particular interest for deployment in resource-constrained environments.

TABLE II
QUALITY OF RECONSTRUCTED IMAGES FOR THE THREE CONFIGURATIONS

Configuration	Accuracy	PSNR	RMSE	UQI	MSE
Reference	97.72%	21.39 dB	0.0852	0.6182	0.0073
Simple Encoder	98.08%	22.33 dB	0.0765	0.6267	0.0059
Simple Decoder	97.50%	21.08 dB	0.0883	0.6054	0.0078

In summary, these models (simplified encoder and decoder models) roughly use only 3% of the computational cost of the



(a) (b) (c)

Fig. 5. Model Loss Observations for Proposed Systems: (a) Reference (Symmetric) Model, (b) Simplified Encoder Model, (c) Simplified Decoder Model.

corresponding reference model, giving away 97% of energy savings in the encoder or decoder. In other words, if the proposed concept is embedded in a WSN or IoT, the battery life of the device can be increased approximately 33 times, which positively contributes to the sustainability of ICT and energy management, as maintenance of these resource constrained devices is very costly financially and environmentally.

The current work is limited to semantic communication based systems implemented using autoencoders and for the specific task of classification of handwritten digits. Further investigation is needed to determine the applicability of this concept to general image processing tasks using autoencoders, as well as on higher resolution images. Exploring the possibility of extending complexity scalability to other common types of deep neural networks used for image related tasks, such as convolutional neural networks and recurrent neural networks, is another area for future research. Deep neural network based video compression, including those based on semantic communications, can also benefit from complexity scalability, especially in simplifying video acquisition on resource constrained devices, such as IoT sensors. Therefore, there is potential for this concept to be applied in a wide range of image and video related applications.

V. Conclusions

We proposed a system based on semantic communication with scalable complexity for image transmission, where the computational complexity of the end-to-end system can be seamlessly shifted to the transmitter (encoder) or the receiver (decoder). Using the MNIST dataset, we demonstrate that this can be done without affecting the overall performance of the communication system. Simulation results suggest that significant energy savings can be made either at the encoder or at the decoder, which will help to prolong battery power up to 33 times, which will be vital in implementing semantic communication-based solutions for resource constrained devices where computationally complex image encoders and decoders are not practically feasible.

These results can be further validated by testing with a larger dataset and can also be extended to video communication systems, allowing for the general adaptation of the proposed asymmetric configurations of semantic communication systems as a sustainable means to address the growing demand for multimedia content. Furthermore, the proposed concept also opens up a novel approach to encoder simplification in media applications attempted by distributed coding, which, in turn, will enable sustainable implementation of video coding systems with high resolution content for multiple applications.

References

[1] M. Landum, M. Moura, and L. Reis, "Ict good practices in alignment with green it," in *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, 2020, pp. 1–6.

[2] M. Azam, H. B. Ameur, Z. Ftiti, A. I. Hunjra, and W. Louhichi, "The role of icts on innovation process of environment and energy convergence," *IEEE Transactions on Engineering Management*, vol. 71, pp. 4418–4440, 2024.

[3] M. K. Banjanin, M. Stojčić, D. Danilović, Z. Ćurguz, M. Vasiljević, and G. Puzić, "Classification and prediction of sustainable quality of experience of telecommunication service users using machine learning models," *Sustainability*, vol. 14, no. 24, 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/24/17053

[4] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[5] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[6] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv e-prints*, p. arXiv:2201.01389, Dec. 2021.

[7] K. Lu, Q. Zhou, R. Li, Z. Zhao, X. Chen, J. Wu, and H. Zhang, "Rethinking modern communication from semantic coding to semantic communication," *IEEE Wireless Communications*, vol. 30, no. 1, pp. 158–164, 2023.

[8] B. Tang, Q. Li, L. Huang, and Y. Yin, "Text semantic communication systems with sentence-level semantic fidelity," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6.

[9] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 55–71, 2023.

[10] Z. Zhang, Q. Yang, S. He, and J. Chen, "Deep learning enabled semantic communication systems for video transmission," in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, 2023, pp. 1–5.

[11] W. Yu and J. Zhao, "Semantic communications, semantic edge computing, and semantic caching with applications to the metaverse and 6g mobile networks," in *2023 IEEE 43rd International Conference on Distributed Computing Systems (ICDCS)*, 2023, pp. 983–984.

[12] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 142–153, 2021.

[13] C. Wang, X. Yu, L. Xu, Z. Wang, and W. Wang, "Multimodal semantic communication accelerated bidirectional caching for 6g mec," *Future Generation Computer Systems*, vol. 140, pp. 225–237, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X22003594

[14] M. U. Lokumarambage, V. S. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," *IEEE Access*, vol. 11, pp. 37 149–37 163, 2023.

[15] J. Chen, G. Bai, S. Liang, and Z. Li, "Automatic image cropping: A computational complexity study," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 507–515.

[16] Y.-L. Tai, S.-J. Huang, C.-C. Chen, and H. H. Lu, "Computational complexity reduction of neural networks of brain tumor image segmentation by introducing fermi-dirac correction functions," *Entropy (Basel, Switzerland)*, vol. 23, no. 2, Feb 11 2021. [Online]. Available: https://www.proquest.com/scholarly-journals/computational-complexity-reduction-neural/docview/2498492460/se-2

[17] M. Díaz, R. Guerra, S. López, and R. Sarmiento, "An algorithm for an accurate detection of anomalies in hyperspectral images with a low computational complexity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1159–1176, 2018.

[18] W. A. C. Fernando, "Special issue on distributed video coding," *Multimedia Tools and Applications*, vol. 48, no. 3, pp. 367–368, 2010. [Online]. Available: https://doi.org/10.1007/s11042-009-0429-y

[19] J. Martínez, W. Weerakkody, W. Fernando, F. Quiles, and A. Kondoz, "Turbo trellis coded modulation for transform domain distributed video coding," *Electronics Letters*, vol. 44, pp. 899–901(2), July 2008.

[20] L. Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[21] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, Y. W. Teh and M. Titterington, Eds., vol. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 249–256. [Online]. Available: https://proceedings.mlr.press/v9/glorot10a.html