

AI-Driven Personalized Remote Education using Adaptive Semantic Communication

Zijun Zhan*, Shaohua Cao[†], Lu Gao[‡], and Zhu Han*

*Electrical and Computer Engineering Department, University of Houston, Houston, TX, USA

[†]Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, China

[‡]Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA

Abstract—With the maturity of communication and multimedia techniques, remote education has become a brand new way of teaching and learning. However, its potential is challenged by limited feedback and lack of customized teaching. Moreover, the knowledge transfer gap is another challenge because different students might prefer distinct learning patterns. To this end, we propose a two-hop semantic communication-empowered remote education framework that can mitigate the knowledge transfer gap while offering a seamless remote education experience. In the proposed framework, we utilize AI-driven adaptive semantic communication to enable efficient knowledge transmission between the teacher and remote students with variety of backgrounds. Given that the knowledge might be distorted during semantic transmission, we propose a semantic restoration component to restore the semantic knowledge. We also develop a student adapter component to customize the semantic knowledge for specific students, thereby meeting students’ expected learning patterns. Experiments on a 3D educational image dataset collected using Unity demonstrate that our method can efficiently transmit knowledge while providing various personalized knowledge formats for students.

Index Terms—Remote education, semantic communication, semantic restoration, knowledge gap.

I. INTRODUCTION

The emergence of the remote education such as Educational Metaverse (Edu-Metaverse) marks a transformative shift in the learning landscape, enabling an equitable and immersive educational experience. For example, coined from “meta” and “universe,” the Metaverse concept encompasses shared, persistent, and decentralized virtual environments where learners interact through digital avatars, engaging in activities that transcend time and geographical constraints [1], [2]. To this end, education is no longer confined to traditional classrooms, allowing learners to study and experiment remotely. Notably, for rural and underrepresented areas, the remote education holds the promise of addressing educational equity by delivering high-quality and engaging content anytime, anywhere.

Despite the promise of remote education, its deployment faces critical challenges in meeting the stringent requirements of ultra-low latency and massive data transmission. Semantic communication, a key vision of 6G, shifts the focus from bit-level accuracy to transmitting and reconstructing the intended meaning, significantly reducing data volume and enhancing robustness [3]. Recent advances in deep joint source-channel coding (DeepJSCC) have shown that deep neural networks

can directly map images to channel symbols, outperforming separate coding schemes under harsh wireless conditions [4]. Moreover, evolving semantic communication frameworks leverage generative models, such as StyleGAN, diffusion models, and variational autoencoders, to further boost efficiency, compress semantic features, and achieve robust transmission even under severe noise or bandwidth limitations [5]–[7].

Semantic communication is a cornerstone of remote education, ensuring seamless knowledge transfer from the teacher to students. The knowledge transfer gap that prevails in traditional on-site education persists in semantic communication-empowered remote education, albeit in a different form. Concretely, the knowledge transfer in the conventional on-site education is a lossy transmission, i.e., students cannot achieve the expected learning outcomes. Different students prefer diverse pedagogical approaches, which further deteriorates the knowledge transfer. Notably, in semantic communication-enabled remote education, knowledge is transmitted in a compressed semantic format to meet the latency requirements, while noise exists during transmission, thereby creating a knowledge transfer gap. Therefore, *we endeavor to design a semantic communication-empowered remote education framework while taking the knowledge transfer gap mitigation into consideration.*

We present our proposed two-hop semantic communication-empowered remote education framework in Fig. 1, where semantic knowledge is delivered in two hops from the teacher to an AI agent, and then from the AI agent to the different students. In the first stage, the teacher encodes the intended instructional knowledge into semantic data through a semantic encoder. The first hop then takes place as semantic transmission, which is inherently lossy and analogous to traditional on-site instruction, where some knowledge may not be fully conveyed. To mitigate this knowledge loss, we incorporate a semantic restoration module—illustrated in this paper with a denoising conditional UNet—that reconstructs the degraded semantics. The second hop is realized by the student adapter, which tailors the restored semantic knowledge to the characteristics and needs of individual students. Finally, the semantic decoder enables students to retrieve and access the adapted knowledge in a form that aligns with the students’ learning requirements. In summary, we conclude our contributions as follows:

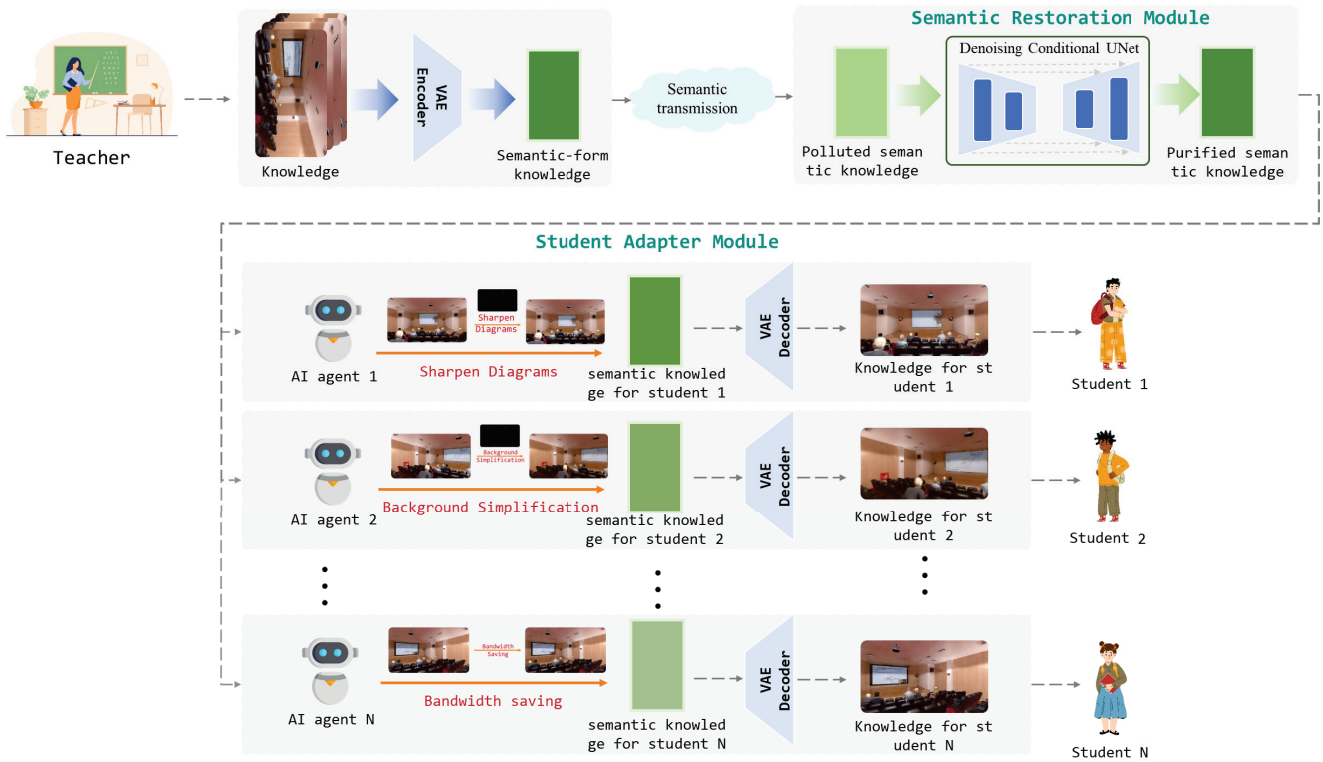


Fig. 1. Illustration of our proposed two-hop semantic communication-empowered remote education framework. In this paper, we consider the knowledge that transfer from the teacher to students is in the format of images. Consequently, we leverage the variational autoencoder (VAE) as the semantic encoder and decoder. Lastly, the semantic restoration module used for lossy semantic knowledge restoration is a denoising conditional UNet.

1. We propose a two-hop semantic communication-empowered remote education framework. We utilize the AI technique to restore eroded knowledge in semantic transmission, while proposing a student adapter to customize the semantic knowledge according to students' diverse characteristics and preferences.
2. We propose a student adapter module and integrate it into the semantic transmission, thereby ensuring the knowledge is customized for students with diverse characteristics and preferences prior to knowledge decoding. Moreover, the kits in the student adapter module are plug-in, which means they can be seamlessly integrated into or detached from the module.
3. We validate the feasibility of our proposed framework using a 3D educational image dataset we collected from a Unity project. Experimental results demonstrate that our proposed framework can restore knowledge lost during semantic transmission while meeting the diverse preferences of students.

The rest of this paper is structured as follows: Section II pinpoints the system framework. We present details of the denoising conditional UNet (an AI technique for restoring lost knowledge) and the student adapter in Section III, and evaluate the proposed approach in Section IV. We also pinpoint several future directions in Section V. Finally, the paper concludes with Section VI.

II. SYSTEM FRAMEWORK AND WORKFLOW

In this section, we outline the key components of our proposed framework, as presented in Section II-A, and describe the framework's workflow in Section II-B.

A. Framework Illustration

As illustrated in Fig. 1, our proposed framework comprises three key modules that facilitate knowledge transfer between the teacher and students. Subsequently, we will illustrate the function and interaction relationships of each module as follows:

Semantic Encoder and Decoder: In semantic communication-empowered remote education, it is necessary for the teacher to compress the knowledge into semantic-form data via a semantic encoder before the knowledge transfer, so as to meet the requirement of remote education. Analogously, before students receive the knowledge, it is vital to utilize the semantic decoder to decode the semantic knowledge.

Semantic Restoration: Considering that the semantic-form knowledge transmission from the teacher to students is an information loss process, we utilize the AI technique to restore the semantic knowledge. In this paper, we consider the knowledge in the form of images. Therefore, we propose using a denoising conditional UNet model to restore semantic knowledge. We illustrate the details of the model in Section III-A.

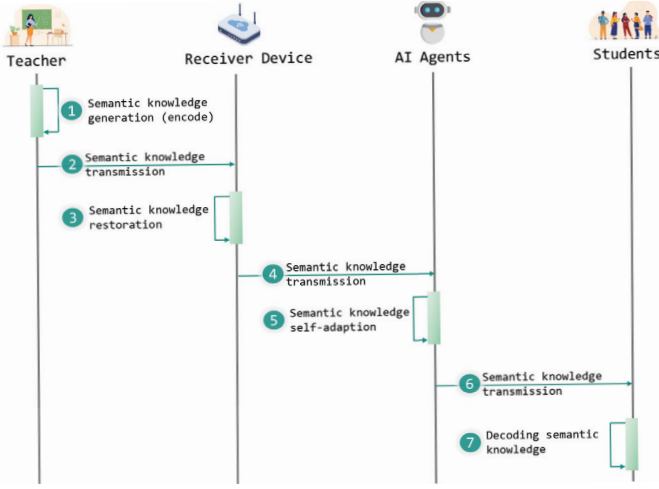


Fig. 2. Workflow diagram of the proposed two-hop semantic communication-empowered remote education framework.

Student Adapter: The restored semantic knowledge can only ensure that students receive the data, but cannot guarantee that students receive the knowledge that the teacher intends to instruct. Therefore, we propose a student adapter that works on semantic data, which can transform the semantic knowledge the teacher intends to instruct into formats that students can grasp with ease. We detail the student adapter in Section III-B.

B. Framework Workflow

We present the workflow diagram of our proposed two-hop semantic communication-empowered remote education framework in Fig. 2, which includes seven steps. Firstly, the teacher utilizes the semantic encoder (VAE encode in this paper) to encode the knowledge (images in this paper) into semantic-form knowledge. Secondly, the teacher transmits the semantic knowledge to students' devices (represented by the receiver device in this paper) via a wireless channel. Thirdly, given that the channel noise will affect the semantic knowledge, an AI technique (a denoising conditional UNet model for noise-polluted image restoration) is implemented to restore the semantic knowledge. Fourthly, the receiver device redirects the restored semantic knowledge to AI agents for knowledge self-adaptation. Fifthly, in light of different students preferring different learning patterns, a semantic knowledge self-adaptation module, a.k.a. student adapter, is integrated to transform the semantic knowledge personalized for each student. Sixthly, the AI agents transmit the adapted semantic knowledge to students. Lastly, students utilize the semantic decoder (VAE decoder in this paper) to decode semantic knowledge and learn customized knowledge with ease.

III. TECHNICAL DETAILS OF PROPOSED FRAMEWORK

We commence with the introduction of the denoising conditional UNet model in Section III-A and then follow with the student adapter in Section III-B.

A. Denoising Conditional UNet Model

We assume the knowledge that the teacher instructs students is in the format of images. We adopt the AI technique of a denoising conditional UNet model to restore the semantic-form knowledge after transmission. It is worth noting that we can adopt different AI techniques for various knowledge formats. For instance, we can use LLM to restore the semantic knowledge for text-form knowledge [8]. Subsequently, we will present details of the denoising conditional UNet model for semantic knowledge restoration.

The kernel of the denoising conditional UNet model is a conditional diffusion model [9] [10]. The diffusion model is primarily composed of forward process and reverse process. The forward process will go through T times, and add a tiny noise into the original image at each time step. Mathematically, the final result of forward process $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$ is derived via

$$q(\mathbf{z}_T|\mathbf{z}_0) = \prod_{t=1}^T q(\mathbf{z}_t|\mathbf{z}_{t-1}) \quad (1)$$

and

$$q(\mathbf{z}_t|\mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t; \sqrt{1 - \beta_t}\mathbf{z}_{t-1}, \beta_t\mathbf{I}). \quad (2)$$

Here, we consider $\mathbf{z} = \mathbf{z}_0$, which is the semantic-form knowledge transmitted by the teacher. β_t is the schedule parameter of the diffusion model.

The reverse process aims to deduce the approximated semantic knowledge $\tilde{\mathbf{z}}_0 = \tilde{\mathbf{z}}$ from the noise sample $\tilde{\mathbf{z}}_T \sim \mathcal{N}(0, \mathbf{I})$. i.e.,

$$p_\phi(\tilde{\mathbf{z}}_0) = p(\tilde{\mathbf{z}}_T) \prod_{t=1}^T p_\phi(\tilde{\mathbf{z}}_{t-1}|\tilde{\mathbf{z}}_t), \quad (3)$$

where $p(\tilde{\mathbf{z}}_T) = \mathcal{N}(0, \mathbf{I})$.

When we shift to the conditional diffusion model, we rewrite (3) as

$$p_\phi(\tilde{\mathbf{z}}_0) = p(\tilde{\mathbf{z}}_T) \prod_{t=1}^T p_\phi(\tilde{\mathbf{z}}_{t-1}|\tilde{\mathbf{z}}_t, \mathbf{z}'), \quad (4)$$

with a conditional input \mathbf{z}' . In this paper, we consider \mathbf{z}' as the semantic knowledge that needs to be restored. Here, $p_\phi(\tilde{\mathbf{z}}_{t-1}|\tilde{\mathbf{z}}_t, \mathbf{z}')$ is similar to (2), we define it as

$$p_\phi(\tilde{\mathbf{z}}_{t-1}|\tilde{\mathbf{z}}_t, \mathbf{z}') = \mathcal{N}(\tilde{\mathbf{z}}_{t-1}; \mu_\phi(\tilde{\mathbf{z}}_t, t, \mathbf{z}'), \sigma_t^2\mathbf{I}), \quad (5)$$

where $\mu_\phi(\tilde{\mathbf{z}}_t, t, \mathbf{z}')$ and σ_t^2 are approximated values of the noise added in the t -th step of the forward process and variance parameter, respectively. We define $\mu_\phi(\tilde{\mathbf{z}}_t, t, \mathbf{z}')$ as

$$\mu_\phi(\tilde{\mathbf{z}}_t, t, \mathbf{z}') = \frac{1}{\sqrt{\alpha_t}}(\tilde{\mathbf{z}}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}}\epsilon_\phi(\tilde{\mathbf{z}}_t, t, \mathbf{z}')), \quad (6)$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{1}^t \alpha_t$. The kernel of the diffusion model is to train a noise approximation function $\epsilon_\phi(\tilde{\mathbf{z}}_t, t, \mathbf{z}')$, which is represented by a denoising conditional UNet model [11].

Given that the conditional diffusion model operates at the semantic knowledge level, it is crucial for us to extract a

TABLE I
UNET ARCHITECTURE STRUCTURE

Input: $\mathbf{z} \in \mathbb{R}^{C' \times H' \times W'}$
Initial Conv2D: $\mathbb{R}^{C' \times H' \times W'} \rightarrow \mathbb{R}^{64 \times H' \times W'}$
Downsampling Path: 4 downsampling blocks, each is defined as: $2 \times \text{ResBlock} + \text{Downsample}$
Bottleneck: $2 \times \text{ResBlock}$
Upsampling Path: 4 upsampling blocks, each is defined as : $\text{Upsample} + (\text{Concat skip}) + 2 \times \text{ResBlock}$
Final Conv2D: $\mathbb{R}^{64 \times H' \times W'} \rightarrow \mathbb{R}^{C' \times H' \times W'}$

corresponding semantic-level dataset from the original teacher knowledge (image dataset). Specifically, we begin by utilizing the semantic encoder (VAE encoder in this paper) to extract the semantic-form knowledge \mathbf{z} from the teacher’s instructed knowledge \mathbf{x} . Next, we simulate the semantic knowledge transmission process on \mathbf{z} and obtain the corrupted semantic knowledge \mathbf{z}' . Repeat the preceding two steps on batch images, and we can obtain a paired dataset $\mathcal{Z} = \{(\mathbf{z}'_i, \mathbf{z}_i) | i \in \{1, \dots, I\}\}$. Notably, the semantic encoder and decoder utilized in this paper are a well-trained version from Stability AI; the parameters of the semantic encoder and decoder are frozen during the conditional diffusion model training. We define the loss function of the model as

$$\ell(\phi) = \mathbb{E}_{\mathbf{z}', \mathbf{z} \sim \mathcal{Z}} [\|\tilde{\mathbf{z}}_0 - \mathbf{z}\|], \quad (7)$$

where $\tilde{\mathbf{z}}_0$ is the restored semantic knowledge of the conditional input \mathbf{z}' . In this manner, we can minimize the difference between the original semantic knowledge \mathbf{z} and the restored semantic knowledge $\tilde{\mathbf{z}}_0$.

The loss function in (7) is applied on $\epsilon_\phi(\tilde{\mathbf{z}}_t, t, \mathbf{z}')$. We present the model architecture of it in Table I for clarity.

B. Student Adapter

Upon restoring the semantic knowledge, we will further leverage a plug-in student adapter to transform the semantic knowledge into various forms tailored for specific students. The student adapter is a collection of AI tools that can transform semantic knowledge to meet specific requirements. In this paper, we primarily consider three AI tools in the student adapter: sharpen diagrams, background simplification, and bandwidth saving.

Sharpen Diagrams: This student adapter will compute the edge map E of decoded semantic knowledge. Next, we downsample the edge map to the size of semantic knowledge and obtain \hat{E} . Subsequently, we utilize \hat{E} to highlight the important areas of the semantic knowledge, and then acquire the transformed semantic knowledge. With this student adapter, students with poor concentration can enjoy a better learning experience.

Background Simplification: This student adapter is similar to the sharpen diagrams, which also require deriving a semantic knowledge size edge map \hat{E} . The difference is that we lower the weight of unimportant areas in the semantic

knowledge as per \hat{E} . This AI tool can offer an alternative option for students with poor concentration, as the factors contributing to low concentration are varied.

Bandwidth Saving: This student adapter leverages the quantization method to compress the semantic knowledge while not affecting the knowledge received by students. This student adapter will be effective when hundreds or even thousands of students are online simultaneously.

Remark: The student adapter is not limited to the above-mentioned three; it can be easily expanded further. For instance, a transformer-based student adapter can help students promptly grasp significant knowledge. Additionally, identifying students’ preferred learning patterns is a crucial prerequisite for adapting the student adapter module. In this paper, we consider two approaches that can be utilized to identify students’ preferred learning patterns. Firstly, we can analyze the interaction logs between students and the AI teaching assistant to predict students’ preferred learning patterns. Secondly, prior to the start of our proposed remote education framework, the teacher can list all of the learning patterns for student selection.

IV. EXPERIMENTAL EVALUATION

In this section, we commence with the introduction of the experimental configurations in Section IV-A. Next, we assess the effectiveness of the semantic restoration component in Section IV-B. Lastly, we assess the effectiveness of the student adapter in Section IV-C.

A. Experimental Configurations

In this paper, we collect 601 images from our developed 3D education Unity project, each image with the size of $3 \times 400 \times 712$. We utilize 80% of the data to train the denoising conditional UNet model and the remaining 20% to evaluate the performance of our proposed framework. For the semantic encoder and decoder, we utilize the Stability AI KL-regularized VAE fine-tuned with MSE loss (“sd-vae-ft-mse”) to encode and decode the knowledge (image data). After encoding, we can obtain the semantic-form knowledge in the shape of $4 \times 50 \times 89$. The compression ratio of our proposed multi-hop semantic communication is 1/48. We consider the signal-to-noise ratio (SNR) in the semantic transmission process to be 18 dB by default.

We evaluate the effectiveness of our proposed framework in two steps. Firstly, we will assess the semantic restoration component via the metrics of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). To assess the effectiveness of the semantic restoration component, we compare it with the following benchmarks:

- 1) **DNSC** [7]: This method uses a vanilla diffusion model to restore the semantic knowledge.
- 2) **DeepSC** [12]: This baseline adopts a 3-layer deep learning convolution neural network to restore the semantic knowledge.

For clarity, we refer to our proposed semantic restoration module as “Ours” in the remainder of this paper. Subsequently, we

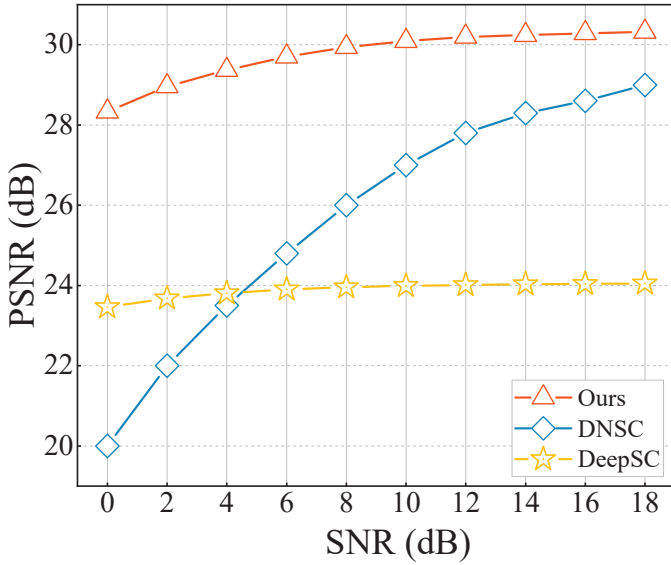


Fig. 3. The PSNR comparison results of the semantic restoration component versus varying SNR in the AWGN channel.

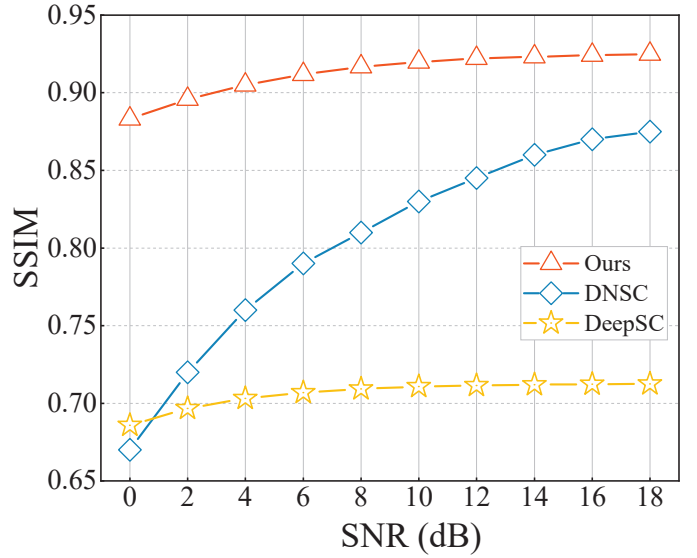


Fig. 4. The SSIM comparison results of the semantic restoration component versus varying SNR in the AWGN channel.

evaluate the student adapter component to determine whether it achieves its intended function.

B. Effectiveness of Semantic Restoration

To assess the effectiveness of the semantic restoration component, we vary the SNR in the semantic transmission process and assess the semantic knowledge restoration quality via metrics of PSNR and SSIM. We vary the SNR in the range of $[0, 18, 2]$ dB and present the semantic restoration results in Figs. 3 and 4.

Observing Figs. 3 and 4, the performance of the semantic restoration component is negatively impacted by the SNR. PSNR and SSIM both increase along with the SNR, which is intuitive. A higher SNR indicates a better semantic transmission environment, as the semantic knowledge loss during transmission will be less; thereby, the difficulty of semantic restoration is attenuated. Notably, the performance of the semantic restoration component can maintain $\text{PSNR} \geq 28$ dB and $\text{SSIM} \geq 0.88$ even in a poor semantic transmission environment with $\text{SNR} = 0$ dB. Moreover, our proposed semantic restoration module consistently outperforms the benchmarks of DNSC and DeepSC even under the same configurations, which demonstrates the effectiveness of the semantic restoration component.

C. Effectiveness of Student Adapter

To evaluate the effectiveness of the three student adapters proposed in Section III-B, we present the results in Figs. 5, 6, and 7, respectively.

Observing the results of the sharpened diagrams student adapter in Fig. 5, we can see that the decoded images are indeed sharpened, especially in the regions labeled with red boxes. Therefore, we can demonstrate the effectiveness of the sharp diagrams student adapter. Similarly, the results

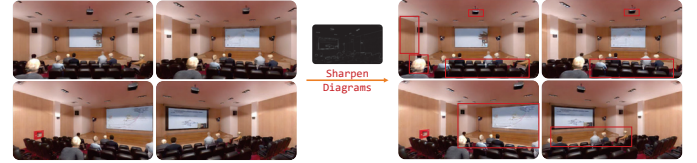


Fig. 5. Effectiveness of the sharp diagrams student adapter. For clarity, we highlight the sharpened areas with red boxes. The left figures represent decoded semantic knowledge after semantic restoration, and the right figures represent decoded semantic knowledge after the student adapter module.



Fig. 6. Effectiveness of the background simplification student adapter. The left figures represent decoded semantic knowledge after semantic restoration, and the right figures represent decoded semantic knowledge after the student adapter module.

in Fig. 6 also validate the effectiveness of the background simplification student adapter, as the chairs in the right images appear much dimmer and blurrier than their left counterparts. Lastly, the results in Fig. 7 can demonstrate the effectiveness of the bandwidth-saving student adapter as well. Concretely, the images after quantization in the right part are visually indistinguishable from the left part. Each element in the original semantic knowledge is a float32, which accounts for 32 bits; however, each element only accounts for 6 bits after adopting the 64-level quantization student adapter.

V. FUTURE WORK

Building upon the proposed two-hop semantic communication-empowered remote education framework, several promis-



Fig. 7. Effectiveness of the bandwidth saving student adapter. We configure 64 quantization levels for this experiment. The left figures represent decoded semantic knowledge after semantic restoration, and the right figures represent decoded semantic knowledge after the student adapter module.

ing research directions warrant further investigation.

First, future research can broaden the scope beyond image-based knowledge to include multi-format semantic content such as text, video, and audio. For example, incorporating large language models (LLMs) for text semantic restoration could substantially improve the general applicability and robustness of the framework across a wide range of educational resources.

Second, to examine system robustness at scale, future work can simulate ultra-large concurrent learning environments where hundreds of students interact with the AI agent simultaneously. To this end, we can inject scalability and universality into our proposed framework.

Finally, since remote education systems often operate under variable network conditions, conducting comprehensive latency and robustness evaluations is a crucial next step. These evaluations should explore different channel qualities, bandwidth constraints, and mobile-device environments, while also considering interoperability with mainstream remote-education platforms.

Collectively, these avenues will advance the proposed framework toward a more scalable, multimodal, and user-adaptive intelligent education system.

VI. CONCLUSION

We proposed a two-hop semantic communication-empowered remote education framework. In this framework, we proposed a semantic restoration component used to restore lost semantic knowledge during semantic knowledge transmission. Moreover, considering that different students might prefer different learning patterns, we further proposed a student adapter component subsequent to the semantic restoration. The student component can transform the semantic knowledge customized for specific students, thereby facilitating students' grasp of the knowledge. Experimental results demonstrated that the semantic restoration component can maintain a PSNR of ≥ 28 dB and an SSIM of ≥ 0.88 under a poor semantic transmission environment. The experimental results also demonstrated that the student adapter component module can achieve the specific expected functions.

ACKNOWLEDGMENT

This work is partially supported by the U.S. National Science Foundation (NSF) ECCS-2302469, Amazon, and Japan

Science and Technology Agency (JST), Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE) JPMJAP2326.

REFERENCES

- [1] H. Lin, S. Wan, W. Gan, J. Chen, and H.-C. Chao, "Metaverse in education: Vision, opportunities, and challenges," in *IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, Dec. 2022.
- [2] X. Chen, D. Zou, H. Xie, and F. L. Wang, "Metaverse in education: Contributors, cooperations, and research themes," *IEEE Transactions on Learning Technologies*, vol. 16, no. 6, pp. 1111–1129, Dec. 2023.
- [3] G. Zheng, Q. Ni, A. Kaushik, and L. Yang, "Semantic communication meets heterogeneous network: Emerging trends, opportunities, and challenges," *arXiv preprint arXiv:2502.08999*, Feb. 2025.
- [4] E. Boursoulatzé, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, Sep. 2019.
- [5] S. Tang, Q. Yang, D. Gündüz, and Z. Zhang, "Evolving semantic communication with generative modelling," in *IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Valencia, Spain, Sep. 2024.
- [6] L. Guo, W. Chen, Y. Sun, B. Ai, N. Pappas, and T. Quek, "Diffusion-driven semantic communication for generative models with bandwidth constraints," *IEEE Transactions on Wireless Communications*, 2025, early Access.
- [7] B. Xu, R. Meng, Y. Chen, X. Xu, C. Dong, and H. Sun, "Latent semantic diffusion-based channel adaptive de-noising semcom for future 6g systems," in *IEEE Global Communications Conference (GLOBECOM)*, Kuala Lumpur, Malaysia, Dec. 2023.
- [8] S. Xu, Z. Wu, H. Zhao, P. Shu, Z. Liu, W. Liao, S. Li, A. Sikora, T. Liu, and X. Li, "Reasoning before comparison: Llm-enhanced semantic similarity metrics for domain specialized text analysis," *arXiv preprint arXiv:2402.11398*, Feb. 2024.
- [9] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, Dec. 2020.
- [10] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *9th International Conference on Learning Representations (ICLR)*, virtual-only, May 2021.
- [11] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF international conference on computer vision (CVPR)*, Vancouver, Canada, Jun. 2023.
- [12] G. Zhang, Q. Hu, Y. Cai, and G. Yu, "Alleviating distortion accumulation in multi-hop semantic communication," *IEEE Communications Letters*, vol. 28, no. 2, pp. 308–312, Feb. 2024.