15TH ITU ACADEMIC CONFERENCE

ITU KALEIDOSCOPE
NEW DELHI 2024

Innovation and digital transformation for a sustainable world

21-23 October 2024
New Delhi, India

Advancing Image Transfer Through Semantic-aided Approaches: A Multi-modal Exploration

22 October 2024

Nargis Fayaz
Ph.D. Scholar, IITD

Session 7 – Enabling technologies

Outline

- **Semantic Communication**
- **Motivation for Advancing Semantic Communication**
- **Challenges in Current Image Transfer Systems**
- **Proposed Solution: Multi-modal Image Transfer**
- **System Overview**
- **Multimodality in Semantic Communication**
- **Comparison of Secondary Mode**
- **Results**
- **Conclusion**

Semantic Communication

Semantic Communication focuses on the **meaning** of the data rather than transmitting raw data

Traditional Communication: Focuses on bit-level accuracy and raw data transfer

Objective: Reduce data transmission size and improve efficiency

"HOW PRECISELY DO THE TRANSMITTED SYMBOLS CONVEY THE DESIRED MEANING?"

THE SEMANTIC PROBLEM

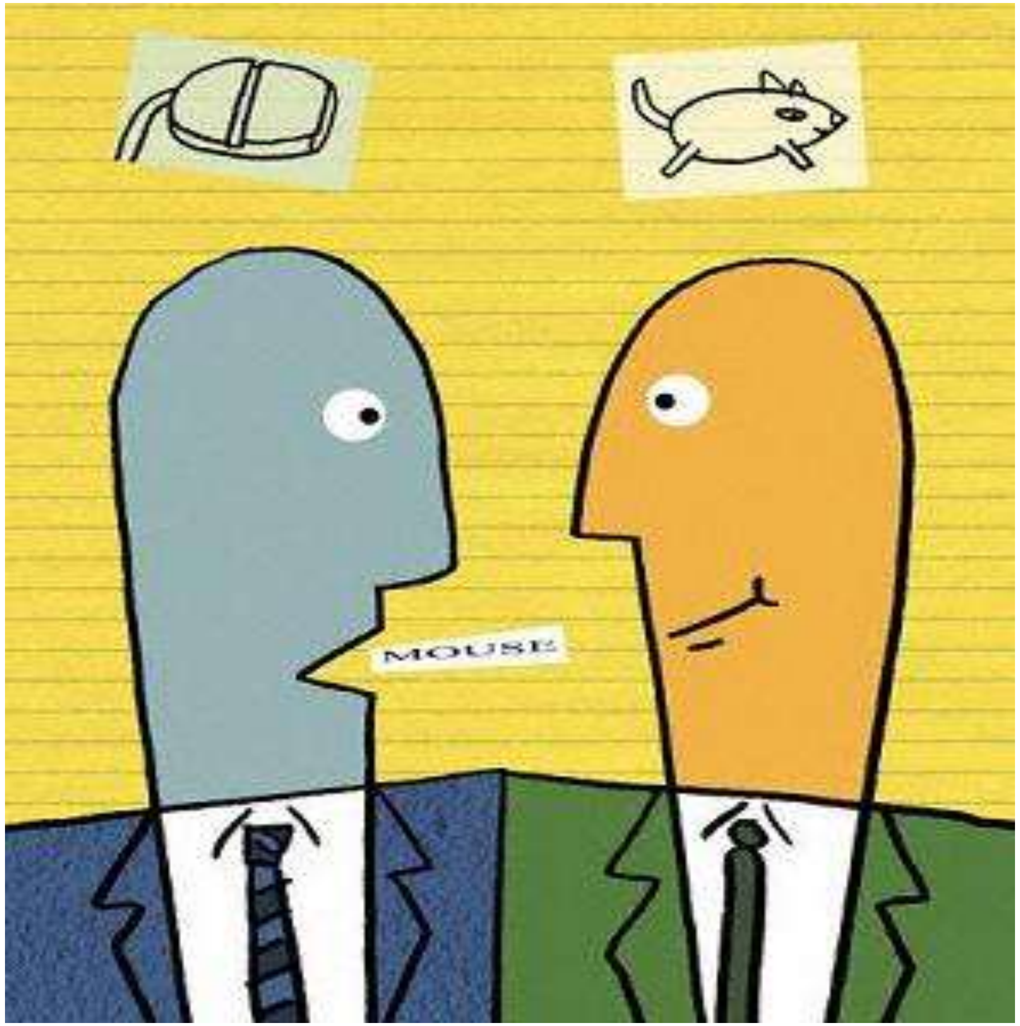


Image Source: Fixing the Economists. (2013, November 5). Is Real Communication Possible? Berkeley's Particularism and Lacan's Semantic Slippage.

Motivation for Advancing Semantic Communication

5G to 6G Transition: Emerging technologies (AI, IoT, autonomous systems) require **higher data rates** and more efficient communication

Challenges in 5G: Latency, **bandwidth constraints**, and lack of real-time adaptability

Solution: SC prioritizes meaning to enhance efficiency, lower latency, and reduce bandwidth consumption

Challenges in Current Image Transfer Systems

Conventional methods transfer raw pixel data, leading to large bandwidth consumption

Single-modality image transmission (e.g., text) fails to capture the full context of images

Need for Innovation: Can we reduce data while maintaining the semantic richness of the image?

Proposed solution: multi-modal image transfer

Introduced a multi-modal approach: Primary Mode: Text captions generated using deep learning models (BLIP)

Secondary Mode: Line art or other structural representations to enhance image fidelity

This approach **balances** data reduction and the preservation of semantic details

→ There is a man holding a lit candle in the desert

→ There is a lizard that is lying down on the ground

Advantages of Image Captioning

- Data Reduction
- Semantic Fidelity

System Architecture Overview

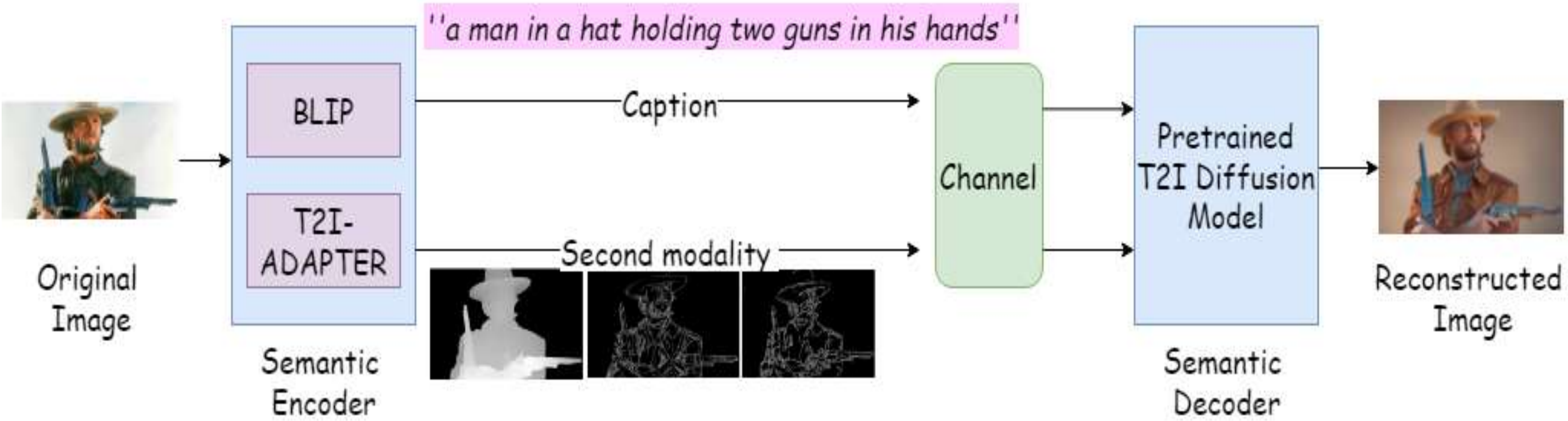


Fig.1 Semantic-aided image transfer through multi-modality

Multi-Modality in Semantic Communication

Why Multi-Modality?

- Single modality (captions only) misses important visual and spatial details
- Multi-modality (captions + structural data) provides a **richer** and **more complete** representation of the image
- Key Benefits:

Increased fidelity with lower data transmission requirements

Better semantic interpretation of images at the receiving end

Comparison of Secondary Modes

- **Depth Map**: Provides spatial information by assigning depth values to pixels
- **Canny Edge**: Focuses on edge detection by locating changes in intensity (used for structural outline)
- **Line Art**: Emphasizes structure and form without color or shading (selected for best balance between fidelity and data reduction)

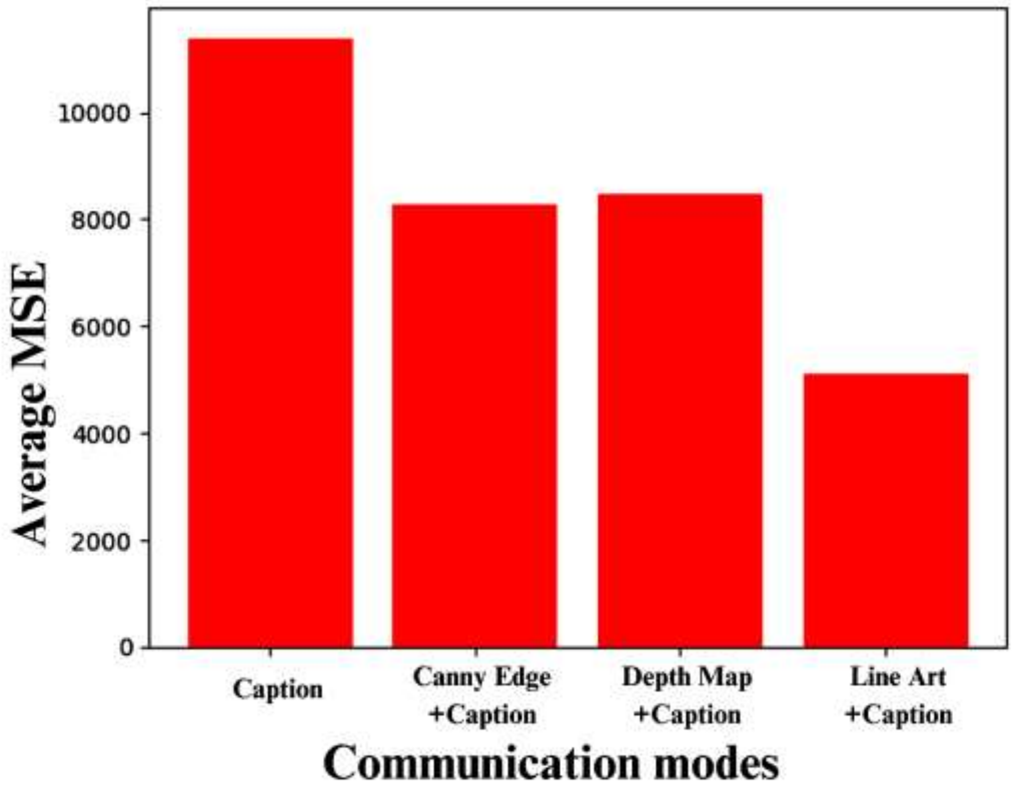
Performance Metrics for Evaluation

- **Mean Squared Error (MSE):** Measures the pixel-level error between the original and reconstructed images
- **Peak Signal-to-Noise Ratio (PSNR):** Indicates the signal quality; higher values represent better reconstructions
- **Structural Similarity Index (SSI):** Evaluates perceptual similarity, focusing on how closely the reconstructed image matches the original from a human visual perspective

Overall Comparison

MODE	ORIGINAL IMAGE	SEMANTIC ENCODING	SEMANTIC DECODING	MSE	SSI	PSNR
LINEART				*7414.3*	*0.6447*	*9.430*
CANNYE--DGE				*1462.4*	*0.4652*	*6.4788*
DEPTHMAP				*15216.3*	*0.4451*	*6.3077*

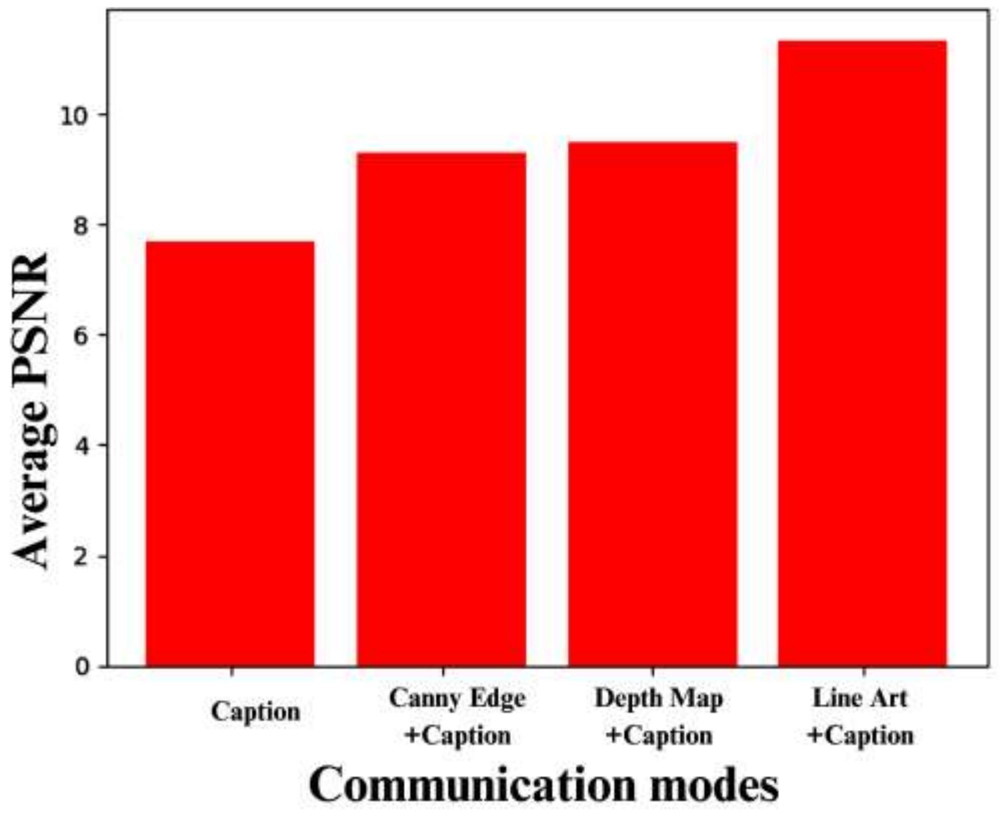
Results: MSE Performance



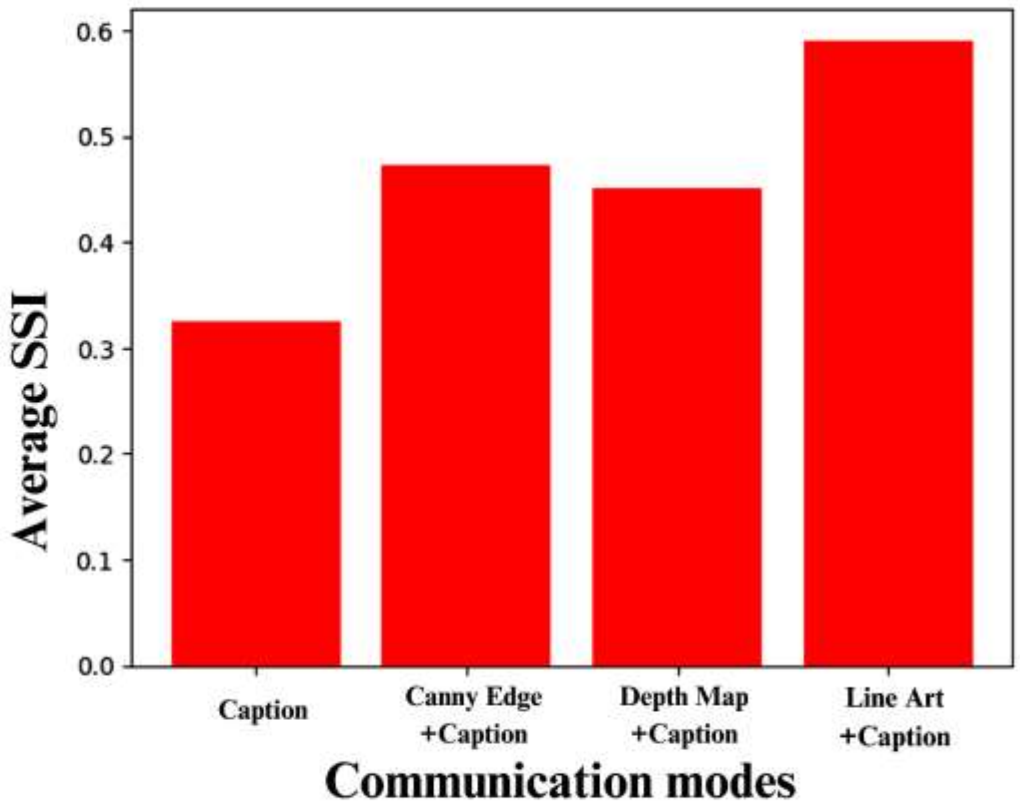
- **Observation**: Lower MSE indicates better image reconstruction accuracy
- **Finding**: Line art consistently shows the lowest MSE, outperforming other modes such as Canny Edge and Depth Map
- **Conclusion**: Line art is the most effective secondary mode for minimizing reconstruction errors

Results: PSNR Performance

- **Observation**: Higher PSNR values reflect better preservation of signal quality
- **Finding**: Line art delivers the highest PSNR among the tested modes, indicating that it maintains the highest fidelity in image reconstruction



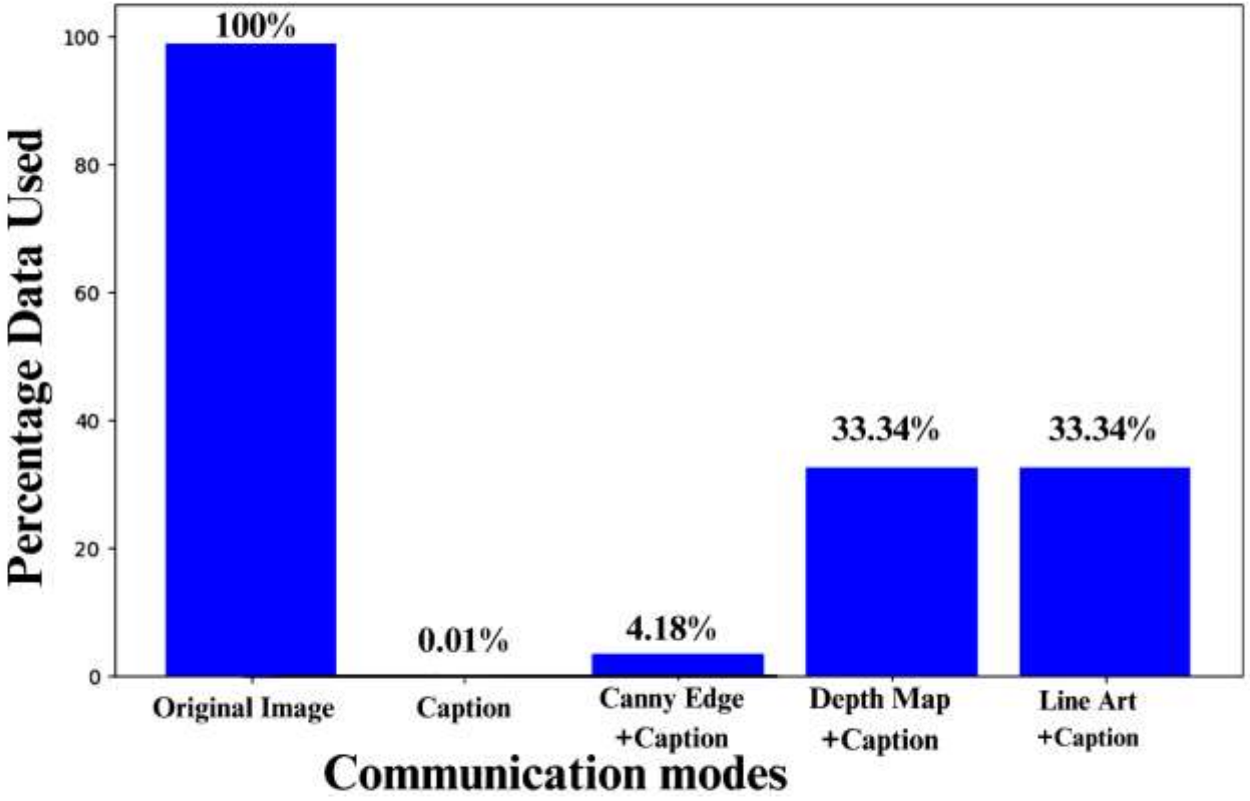
Results: SSI Performance



- **Observation**: SSI measures the visual similarity of reconstructed images to the original
- **Finding**: Line art shows the highest SSI, making it the most effective at producing images that are perceptually similar to the original

Data Reduction Analysis

- **Original Image**: 728x492 pixels, requiring 8.59 million bits
- **Caption + Line Art** Significant data reduction, with line art requiring 2.86 million bits and the caption requiring 312 bits
- **Conclusion**: This multi-modal approach reduces the data required by nearly **65%** while maintaining image fidelity



Conclusion

- **Summary**:
    - Proposed a novel **multi-modal system** for semantic-aided image transfer using **captions and line art**
    - Demonstrated significant **data reduction** with minimal loss in image quality
    - **Line art** emerges as the optimal second mode for preserving image structure and minimizing data
- **Future Work**:
    - Investigate the addition of **color consistency** to further improve reconstruction accuracy
    - Explore more complex images and **dynamic content** in real-time applications

References

[1] X. Luo, H.-H. Chen, and Q. Guo, “Semantic communications: Overview, open issues, and future research directions,” IEEE Wireless Communications, Vol. 29, no. 1, pp. 210–219, 2022.

[2] 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, Jan. 2023.

[3] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong,D. Niyato, X. Chi, X. Shen, and C. Miao, “Semantic communications for future internet: Fundamentals, applications, and challenges,” IEEE Communications Surveys Tutorials, vol. 25, no. 1, pp. 213–250, 2023.

[4] G. Yin, B. Liu, L. Sheng, N. Yu, X. Wang, and J. Shao, "Semantics disentangling for text-to-image generation,” in Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, 2019, pp.2327–2336.

[5] 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, 2023.

[6] J. Li, D. Li, C. Xiong, and S. Hoi, “Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation,” International conference on machine learning. PMLR,2022, pp. 12 888–12 900.

[7] C. Mou, X. Wang, L. Xie, Y. Wu, J. Zhang, Z. Qi, and Y. Shan, “T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 5, 2024, pp.4296–4304.[

8] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, and H. Laga, “A comprehensive survey of deep learning for image captioning,” ACM Computing Surveys (CsUR), vol. 51,no. 6, pp. 1–36, 2019.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P.Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.

[10] A. Hertzmann, “Why do line drawings work? a realism hypothesis,” Journal of Vision, vol. 21, no. 9, pp.2029–2029, 2021.

Thank you for your attention.
Any questions?

Authors' emails: Dawood Aziz Zargar dawoodaziz_2021bece016@nitsri.ac.in
Hashim Aijaz hashim_2021bece007@nitsri.ac.in
Nargis Fayaz eez218533@ee.iitd.ac.in

Thank you!