LMM-driven Semantic Image-Text Coding for Ultra-low Bitrate Learned Image Compression

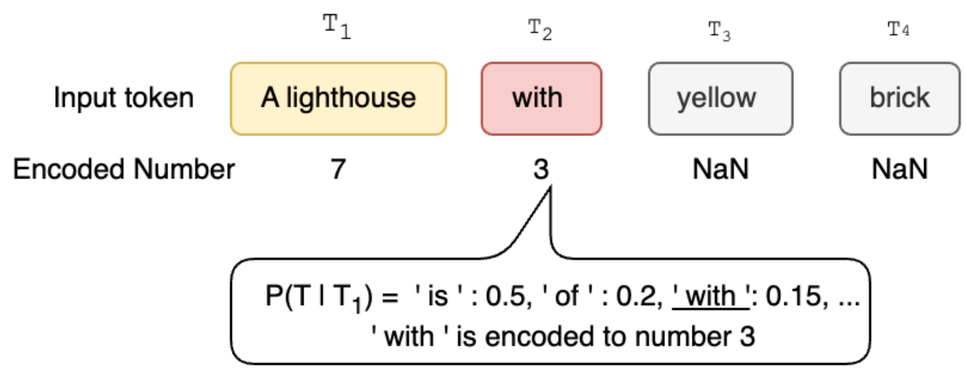
Shimon Murai, Heming Sun, Jiro Katto
Department of Computer Science and Communications Engineering, Waseda University, Tokyo, Japan
Faculty of Engineering, Yokohama National University, Kanagawa, Japan

Background

- Learned Image Compression is a technique to design neural network-based non-linear transform and entropy model to compress images

- Its sub-domain, **ultra low-bitrate compression**, targets to less than **0.1 bpp** while keeping good perceptual quality
- Some models utilize **text caption** as a sub-information that captures semantic information (MISC[3], Text & Sketch[4])

- Recently, **Large Language Model** is shown to be effective in text compression and data compression (LLM-ZIP[1], Language Modeling is Compression[2])

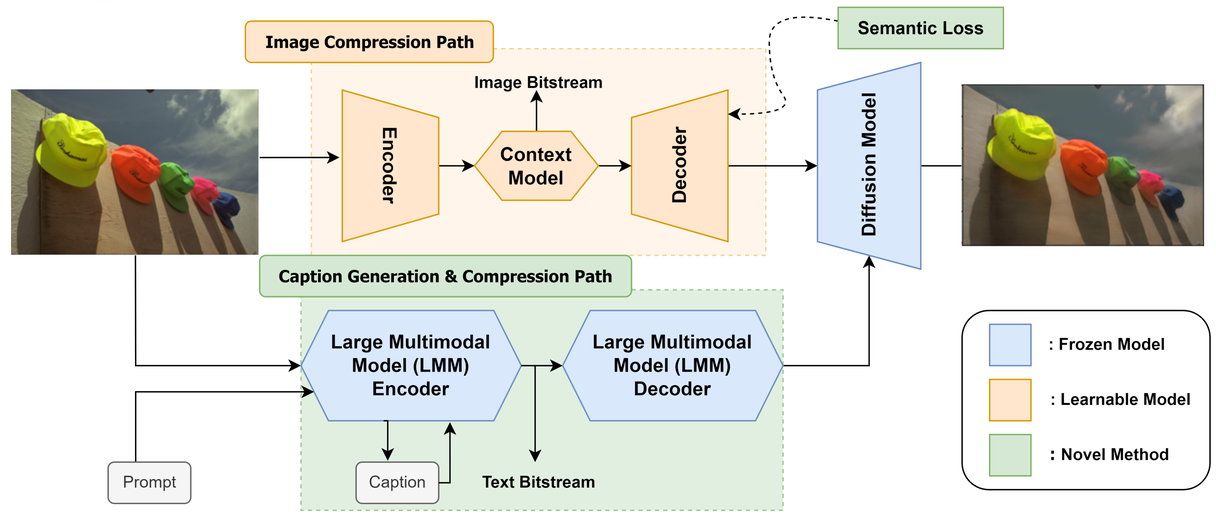


- How can we integrate **LLM based compression** in the workflow of **Learned Image Compression?**

Methodology

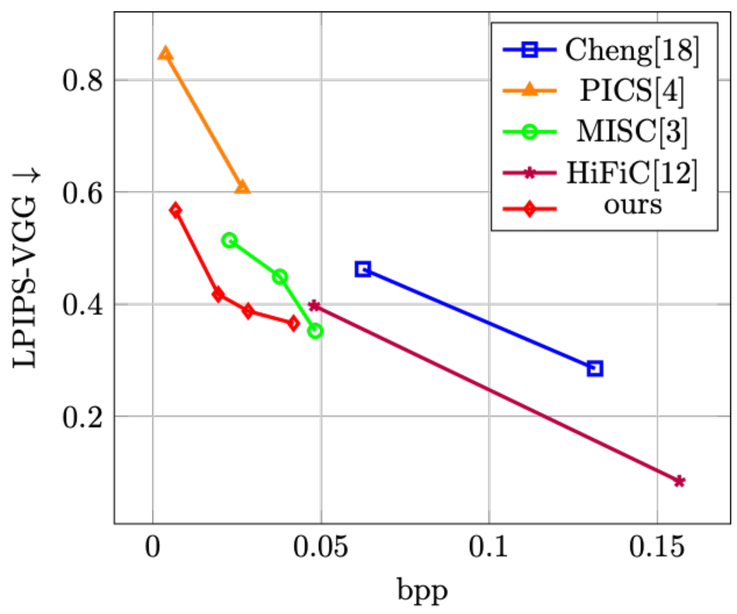
- We generate text caption, and compress it into bitstream **with one large multimodal model**
- Input images are fed to fine-tuned **low-bitrate image compressor**
- Output (distorted) images are then refined with **generative model** conditioned with the caption

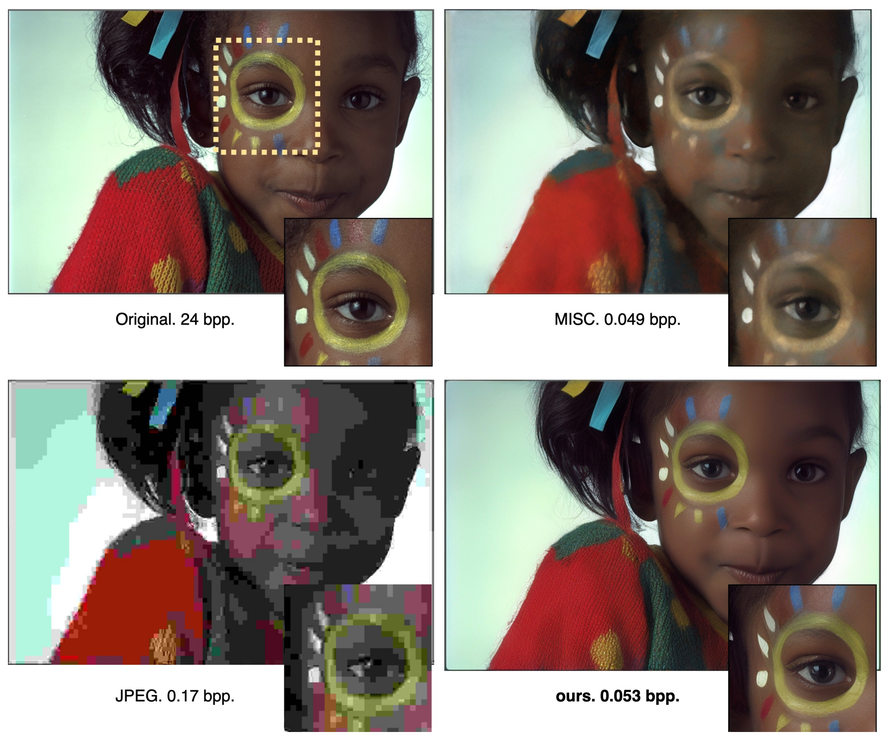
- Our contributions are:
  - We show that **captioning and its compression** can be done in **one LMM**
  - Developed efficient fine-tune methods with **perceptual and semantic loss**



Results & Takeaways

- Our model **achieves more than 65% text compression ratio** and more than **40% bitrate saving** than existing methods (in LPIPS BD rate)
- Example images show our methods eliminate color distortion under ultra low-bitrate (0.053 bpp!)





Code & Demo Available on **GitHub**

[1]C. S. K. Valmeekam, K. Narayanan, D. Kalathil, J.-F. Chamberland,and S. Shakkottai, "LLMZip: Lossless Text Compression using LargeLanguage Models," arXiv preprint arXiv:2306.04050, June 2023.
[2]G. Deletang, A. Ruoss, P.-A. Duquenne, E. Catt, T. Genewein, C. Mat- tern, J. Grau-Moya, L. K. Wenliang, M. Aitchison, L. Orseau, M. Hutter, and J. Veness, "Language modeling is compression," in ICLR, 2024
[3]C. Li, G. Lu, D. Feng, H. Wu, Z. Zhang, X. Liu, G. Zhai, W. Lin, and W. Zhang, "MISC: Ultra-low Bitrate Image Semantic Compression Driven by Large Multimodal Model," arXiv preprint arXiv:2402.16749, no. arXiv:2402.16749, 2024.
[4]E. Lei, Y. B. Uslu, H. Hassani, and S. S. Bidokhti, "Text + sketch: Image compression at ultra low rates," ICML 2023 Workshop Neural Compression, 2023.
This work is supported in part by SCAT, in part by JSPS KAKENHI Grant Number JP23K16861, and in part by Telecommunications Advancement Foundation.