Digital Semantic Communication with Neural Image Compression

Long V. Nguyen*, Tuan T. Nguyen†, Octavia A. Dobre* and Trung Q. Duong*
* Memorial University, Canada, e-mail: {lnguyen, odobre, tduong}@mun.ca
† University of Greenwich, UK, e-mail: tuan.nguyen@greenwich.ac.uk

***Abstract*—Although analog semantic communication systems have attracted significant attention recently, there has been relatively less focus on digital semantic communication systems. In this work, we introduce a neural image compression-enabled semantic communication system to enhance the efficiency of digital image transmission, named NCSC. By employing an accurate and adaptable entropy model, NCSC obtains the efficiently compressed bitstreams, which are compatible with digital communication systems. Incorporating with the well-established digital components, our system trained on the MS-SSIM metric can achieve a significant bandwidth compression ratio of 0.002 at low SNR, reducing remarkably transmission overhead. Extensive simulations show that our approach outperforms traditional digital communication systems in terms of perceptual quality and bandwidth efficiency under challenging channel conditions.**

I. Introduction

The field of communication has undergone a paradigm shift with the recent emergence of semantic communication. Unlike traditional methods that focus solely on the reliable transmission of bits, semantic communication [1] aims to convey the meaning and intent of information, promising enhanced efficiency and robustness in varying environments. While early semantic communication research explored analog semantic transmission schemes, practical deployment necessitates a transition towards digital semantic communication systems to leverage existing infrastructure and benefit from well-established digital signal processing techniques. This transition inherently offers better scalability and flexibility, making them suitable for a wide range of applications, including the rapidly growing domains of IoT, 5G, and beyond.

Neural data compression [2], powered by deep learning architectures, has emerged as a groundbreaking approach in image compression, demonstrating remarkable capabilities in achieving unprecedented compression ratios while maintaining high perceptual quality. Unlike traditional compression methods such as JPEG, JPEG2000, which rely on hand-crafted features and predetermined transformation techniques, neural compression leverages learned representations to capture complex patterns and semantic features within the data. Neural compression techniques can achieve higher compression than conventional methods while maintaining comparable or better image quality. This capability is not only crucial for efficient image storage, but also has profound implications for communication systems, enabling the transmission of richer and more complex data at reduced bandwidth requirements.

This paper introduces a novel digital semantic communication system that integrates the advantages of neural image compression. By employing a neural compression-based entropy coder, our proposed system generates a highly compressed bitstream of the original image for transmission. These compressed representations can be efficiently transmitted and combined with widely adopted digital components in practice. At the receiver, the image is reconstructed from the transmitted bitstreams by the learned decoder. This allows us to achieve highly efficient data compression while maintaining the semantic content of the transmitted information, leading to more efficient and robust communication systems. We demonstrate that this approach can minimize the required transmission overhead significantly while attaining the visually pleasant reconstructed images, providing a pathway towards more efficient and intelligent digital communication systems.

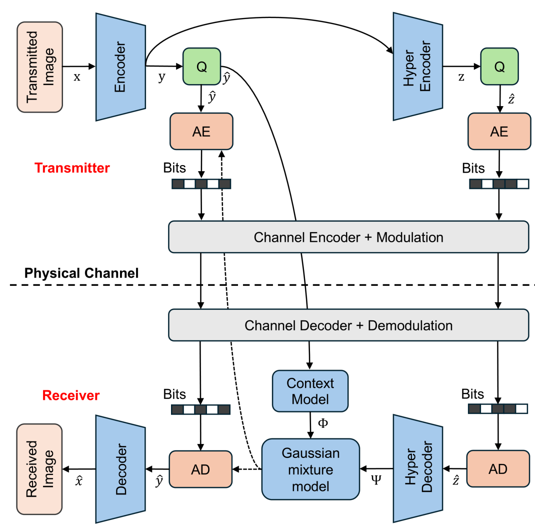

Fig. 1: Architecture of neural compression-based digital semantic communication (NCSC)

II. System Model

At the transmitter, the encoder maps an input RGB image x into a low-dimensional latent representation, followed by a quantization to generate $\hat{y}$. An arithmetic coder is utilized to encode the quantized codes into a bitstream. A similar process is applied to the Hyper Encoder, a subnetwork for learning a probabilistic model over quantized latents used for entropy coding. These compressed bitstreams are then transmitted through the wireless channel, which is compatible