Dresden University of Technology

ComNets
Deutsche Telekom Chair of Communication Networks

Fakultät Elektrotechnik und Informationstechnik

Institut für Nachrichtentechnik, Deutsche Telekom Professur für Kommunikationsnetze

AI-based Semantic Compression for AR/AI Glasses

Collaboration: Ericsson | Deutsche Telekom Chair of Communication Networks
Position: WHK (wissenschaftliche Hilfskraft) or Master's Thesis.

Background:

Augmented Reality (AR) and AI-enabled smart glasses generate continuous streams of high-resolution image data that must be transmitted to remote servers for intelligent processing. Traditional image compression methods (e.g., JPEG/PNG) are not optimized for machine understanding, leading to inefficient network bandwidth use and increased latency in AI-assisted applications. This work focuses on developing and evaluating an AI-based semantic compression pipeline for AR/AI glasses. The core idea is to employ a semantic encoder at the user device that compresses image frames into compact vector codes (VC), which are then transmitted over the network to a remote AI inference server. At the server side, a semantic decoder reconstructs meaningful prompts or keywords from the received vector codes, which are subsequently fed into a Large Language Model (LLM) agent. The LLM agent then generates and returns an audio description of the scene to the user.
The goal is to investigate, implement, and benchmark a lightweight, AI-based semantic compression scheme using open-source datasets and, ultimately, to prototype an end-to-end semantic compression pipeline with real AR glasses hardware.

Tasks:

- Gain an understanding of the fundamentals of semantic communication and AI-based image compression.
- Investigate and compare existing open-source semantic encoder/decoder architectures suitable for edge deployment.
- Set up and train a semantic encoder–decoder pipeline using open-source image datasets.
- Evaluate compression performance in terms of semantic fidelity, compression ratio, and latency.
- Integrate the semantic decoder output with an LLM agent (e.g., via You.com API) for prompt/keyword-based scene description generation.
- Set up the API platform (You.com, TU Dresden) for LLM-based inference.
- (If applicable) Acquire and configure AR glasses hardware and demonstrate the end-to-end semantic compression prototype.
- Document results and provide recommendations for deployment in real-time AR/AI scenarios.

Keywords:

Semantic Compression, AR/AI Glasses, Vector Codes, Semantic Encoder-Decoder, LLM Agent, Edge AI, Image-to-Text

Contact:

- Supervisors: Sifat Rezwan, Dr.-Ing. Juan A. Cabrera
- Language: English