LatentCSI: Real-Time Reconstruction of Physical Scenes from WiFi CSI via Latent Diffusion

Eshan Ramesh
esrh@esrh.me
Institute of Science Tokyo
Tokyo, Japan

Takayuki Nishio
nishio@ict.e.titech.ac.jp
Institute of Science Tokyo
Tokyo, Japan

Abstract

We demonstrate real-time high-resolution generation of images of the physical environment from WiFi CSI. Our demo is based on LatentCSI, our novel CSI-to-image generation framework that encodes CSI samples into the latent space of a pretrained latent diffusion model (LDM) to produce high-quality images with optional editing & reconstruction capability by text prompts. Our prototype consists of sensor nodes to collect ground truth CSI and camera images, an edge and training server to host and update LatentCSI components, and a client server that serves predicted images to clients. Our use of latent space offers faster training and inference at better quality. This allows for online model training at ~30 samples/sec, and inference with a latency of ~60ms. To the best of our knowledge, this work presents the first real-time demonstration of image generation from WiFi CSI.

CCS Concepts

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**.

Keywords

WiFi sensing, CSI sensing, latent diffusion, image generation

ACM Reference Format:
Eshan Ramesh and Takayuki Nishio. 2025. LatentCSI: Real-Time Reconstruction of Physical Scenes from WiFi CSI via Latent Diffusion . In *Proceedings of The 31st Annual International Conference on Mobile Computing and Networking (ACM MobiCom '25).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3680207.3675601

ACM MOBICOM '25, November 4–8, 2025, Hong Kong, China
© 2025 Copyright is held by the owner/author(s).
ACM ISBN 979-8-4007-1129-9/2025/11.
https://doi.org/10.1145/3680207.3765601

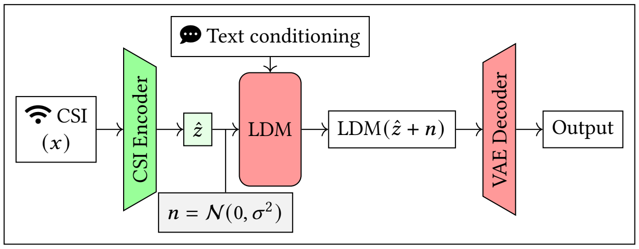


Figure 1: LatentCSI LDM with custom CSI encoder

1 Overview

WiFi channel state information (CSI) has been widely applied to sensing tasks such as pose estimation and activity classification [2]. CSI contains rich information about the surrounding propagation environment by characterizing the channel's effect on various subcarriers. While most existing studies focus on predicting low-dimensional indicators of the physical world (human pose, location, gestures, etc.) research on reconstructing high-dimensional representations remains relatively limited.

We proposed LatentCSI [3], shown in Figure 1, a method that generates high-resolution camera images from WiFi CSI by leveraging a pretrained latent diffusion model (LDM). Unlike computationally intensive prior approaches, LatentCSI replaces the encoder of the LDM with a lightweight neural network to map CSI amplitudes into the latent space of the LDM. The LDM's denoising diffusion model is then applied to the latent vector with text guidance, followed by decoding via the LDM's decoder to obtain a predicted image. This design avoids the overhead of pixel-space image generation, and allows details lost in the dimensional bottleneck to be reconstructed by diffusion models. The target latent vector is approximately 90% smaller than a full-resolution image, enabling faster inference and training while also suggesting a natural method to split processing across several nodes.

2 System Design

Figure 3 illustrates our distributed and scalable architecture, which consists of four loosely coupled nodes: the **Sensor Hub (SH)**, **Edge Server (ES)**, **LatentCSI Server (LS)**, and **Client Server (CS)**.

2.1 Online model training

(1) SH, a lightweight node, captures synchronized RGB images (512×512 image, ~786kB) and WiFi CSI measurements at 30Hz. Camera images are sent to the ES for encoding into ground-truth latent vectors.

(2) ES, a GPU-capable edge node, hosts a pretrained image encoder and uses it to receive images, encodes them into ground-truth latent vectors (4 × 64 × 64 array, ~65kB), and returns them to the SH, which forwards (CSI, latent) pairs to the LS.

(3) LS, a GPU-powerful cloud node, hosts a compact CSI encoder (approximately 20M parameters). The LS updates the model weights using batches of (CSI, latent) pairs received from the SH.

2.2 Real-time image generation from CSI

(1) CS is a web server that concurrently serves multiple clients and hosts a pre-trained image decoder. Upon receiving a request, the CS contacts an SH to retrieve a fresh CSI sample. This is forwarded to the LS along with optional diffusion parameters (e.g. text prompt, modification strength).

(2) LS runs the CSI encoder on the received CSI sample and returns the predicted latent vector. If diffusion parameters were provided, it applies the latent diffusion model to the vector and returns the transformed output.

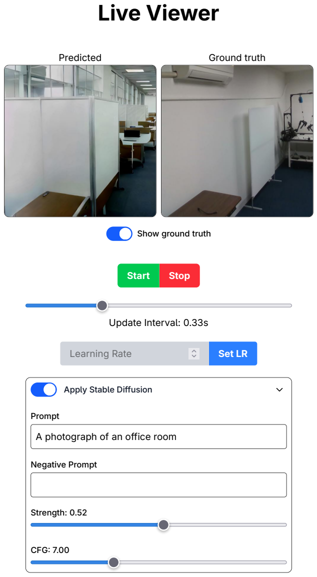
(3) CS decodes the received latent vector using a pre-trained decoder and displays the image to the client through the web server.

3 Demonstration

In our demo, an SH consists of a small NUC computer with an AX210 WiFi chip that supports 160MHz CSI measurements and an Intel RealSense camera for RGB images. The ES is an NVIDIA Jetson Orin AGX, connected via Ethernet (~2.3 Gbps) to the SH to support high-bandwidth image throughput.

The LS is utilizes a powerful GPU but has a comparatively slow connection. By exchanging only small latent vectors instead of full-resolution images, utilized bandwidth between the SH/CS and LS is reduced by over 90%. The decoupled nature of the LS means that it can be scaled according to client requirements and throughput. In our demo, the LS runs a slightly modified model from our LatentCSI work, and uses the Stable Diffusion v1.5 LDM [4].

The CS is located on-site and is responsible for decoding latents from the LS and serving them to clients on a web server, shown in Figure 2. The CS uses TAESD [1], a fast autoencoder to enable high-throughput. End-to-end inference is dominated by latent decoding performance, although running the CS on a modern GPU laptop is sufficient to achieve a 60ms latency. Visitors connect to the CS from their own devices and watch predictions in real time.



Update Interval: 0.33s

Prompt: A photograph of an office room

Negative Prompt:

Strength: 0.52

CFG: 7.00

Figure 2: Client Server web UI

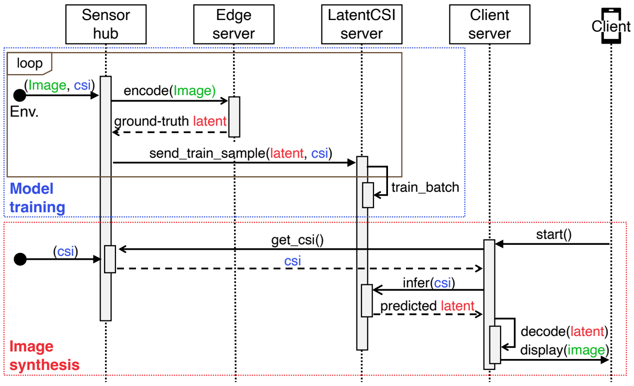
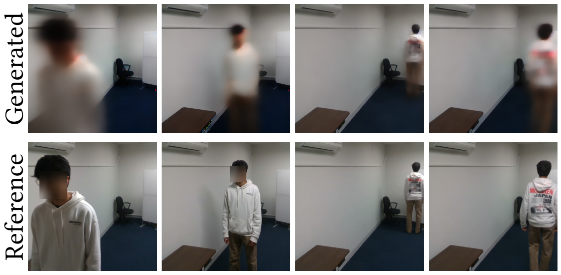


Figure 3: Sequence diagram of the LatentCSI demo.

Table 1: Sample images after 5min of pretraining



Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 23K24831 and 23K26109, and JST ASPIRE Grant Number JPMJAP2346, Japan

References

[1] Ollin Boer Bohan. 2025. Tiny Autoencoder for Stable Diffusion (TAESD). https://github.com/madebyollin/taesd. (2025).

[2] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi Sensing with Channel State Information: A Survey. *ACM Comput. Surv.*, 52, 3, (June 2019), 46:1–46:36.

[3] Eshan Ramesh and Takayuki Nishio. 2025. High-resolution efficient image generation from WiFi CSI using a pretrained latent diffusion model. (July 2025). arXiv: 2506.10605 [cs].

[4] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEECVF Conf. Comput. Vis. Pattern Recognit. CVPR.* (June 2022), 10674–10685.