LatentCSI

Eshan Ramesh, Takayuki Nishio

Institute of Science Tokyo

November 18, 2025

Motivation: High-dimensional sensing with CSI is hard!

There is plenty of research demonstrating that CSI is usable for low-dimensional sensing:

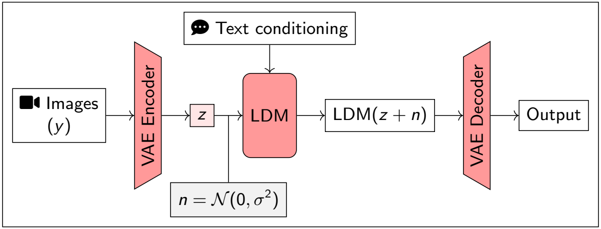
- Localization
- People-counting
- Pose estimation
- etc.

Image generation is useful but hard

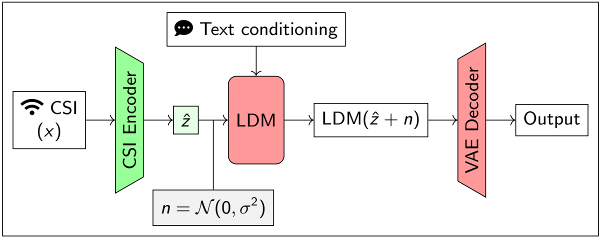
- Target is a $512 \times 512 \times 3$ image $\rightarrow$ 786,432 output values
- Input is nowhere near as large!

Idea: Predict smaller latent embeddings instead of images!

Adapt a latent diffusion image-to-image process:



With a small **CSI encoder** to make it a CSI-to-image pipeline.



LatentCSI Advantages

- Small output allows lighter models that train and infer faster.
- Latent bottleneck loses fine high-dimensional details (faces, text, etc). **This is good!**.
- **Any image autoencoder works!**
  - Using the latent space of an LDM lets us transform images without an extra encoding step.
  - Fine details lost in the latent bottleneck can be controllably simulated.
  - Many autoencoders available for Stable Diffusion: we can choose an AE to balance performance & quality.
- *Operating in latent space also suggests a simple way to distribute training and inference!*

Supervised results

Table: Our in-house dataset (160MHz HE)

Ref	Ours	Base

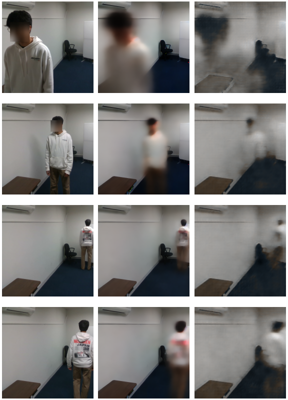
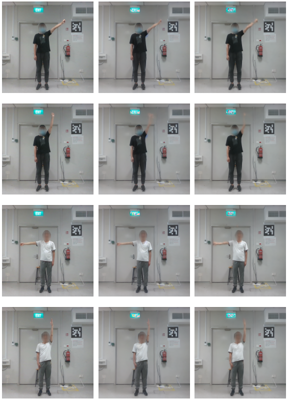


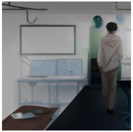
Table: MM-Fi dataset(3×40MHz VHT)

Ref	Ours	Base



Stable diffusion

Table: Text-guided CSI Image Generation

	Reference	"a photograph of a man in a small office room, 4k, realistic"	"a drawing of a man in a laboratory, anime, 4k"
Dataset 1			
Dataset 2			

The choice of latent space as that of a latent diffusion model like Stable Diffusion lets us transform images with a text-prompt efficiently.

Supervised results: quantitative performance

Table: Quantitative comparison of methods over 5 runs

Method	FID ↓	RMSE ↓	FID (crop) ↓	RMSE (crop) ↓	Time (min:s)
Dataset 1 LatentCSI	**134.23 ± 7.18**	**18.95 ± 0.44**	**126.09 ± 7.15**	**19.81 ± 0.66**	**05:02 ± 00:46**
Dataset 1 Baseline	268.03 ± 4.23	20.45 ± 0.27	296.93 ± 4.83	21.74 ± 0.29	16:29 ± 03:35
Dataset 2 LatentCSI	**28.21 ± 0.78**	7.87 ± 0.09	**27.90 ± 0.43**	7.90 ± 0.03	**16:24 ± 00:03**
Dataset 2 Baseline	47.67 ± 2.04	**7.15 ± 0.50**	69.34 ± 2.50	**7.19 ± 0.24**	91:09 ± 19:19

- **LatentCSI outperforms a non-latent baseline of comparable parameter count**
- With rough data quality (dataset 1): LatentCSI has much better fidelity, much faster.
- With good data quality (dataset 2): LatentCSI still has better fidelity even though raw accuracy is lower.
- Bigger gaps when the region containing a person is selected.

Implementation: real time CSI to image

Offline supervised results are good but... is a real-time online system feasible?

Yes!
LatentCSI models can be trained so fast that a real-time implementation is practical

- Natural distributed approach:
  - Sensor(s) collect training data, send to a training server
  - Client(s) connect to the server, make inference requests
- Can put the training server to the cloud and avoid ever transmitting full images (semantic communication!)
  - **Transmission cost reduced by 90%!**
  - Better privacy guarantees with a small-enough latent size

Implementation: Training process

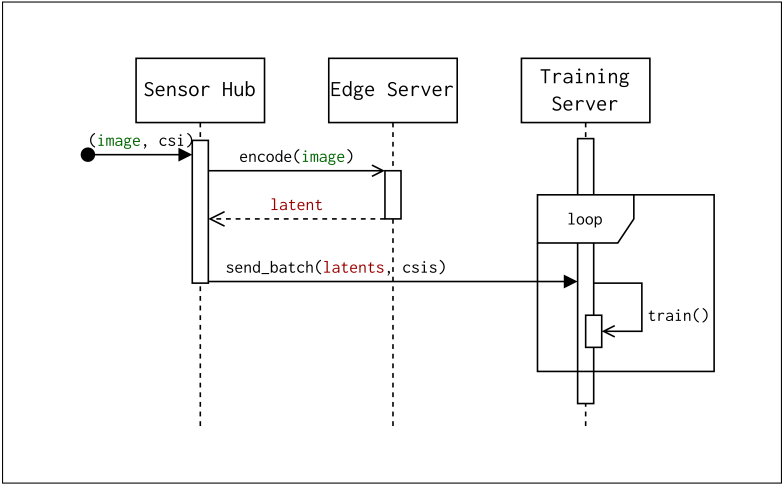


Figure: Training process

Implementation: Inference process

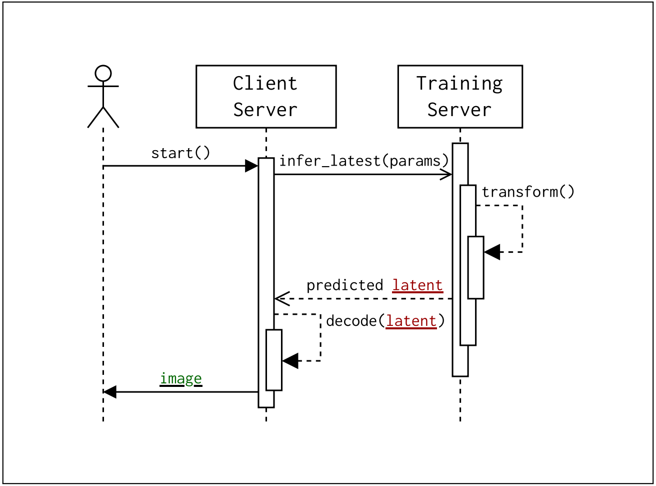
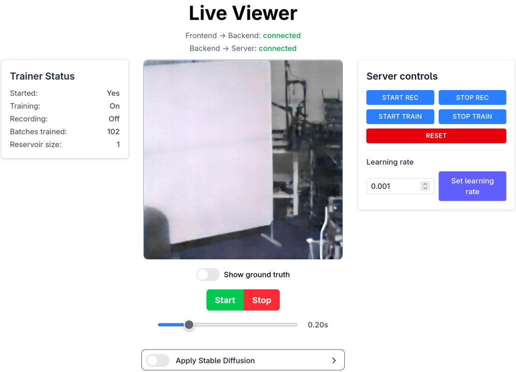


Figure: Inference process

Implementation



- NVIDIA Jetson used for edge latent encoding
- Training server hosted on AWS with a powerful GPU
- **Sensor $\rightarrow$ Server: 60 samples/s, 1s latency**
- **Client $\leftrightarrow$ Server: $\geq$ 60 fps**

Conclusion

Latent methods are powerful for high-dimensional inverse problems

In our application to CSI-to-image generation, they enable:

- High-resolution image generation with fast performance
- **A prototype of real-time image to csi!**

Links

Paper (Supervised results)
`arxiv.org/abs/2506.10605`

Code & Documentation
`github.com/nishio-laboratory/latentcsi`