Assessing Learned Models for Phase-only Hologram Compression

Zicong Peng
University College London
London, UK
zicong.peng.24@ucl.ac.uk

Yicheng Zhan
University College London
London, UK
ucaby83@ucl.ac.uk

Josef Spjut
NVIDIA
Durham, NC, US
jspjut@nvidia.com

Kaan Akşit
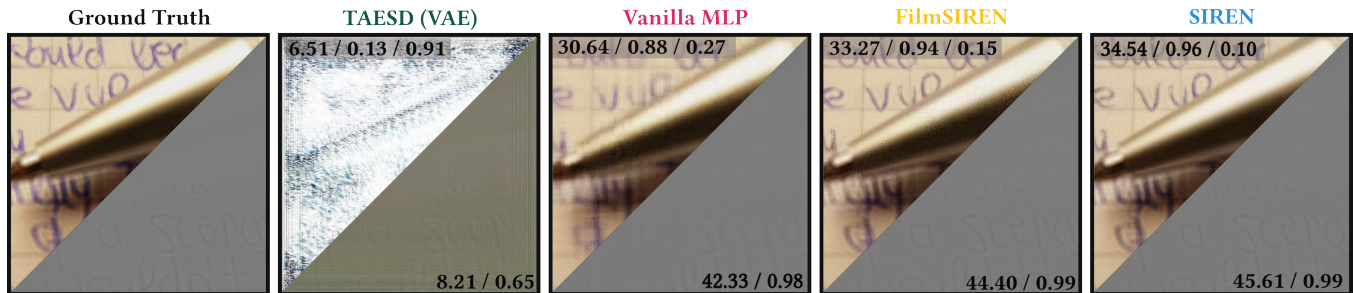University College London
London, UK
k.aksit@ucl.ac.uk

Figure 1: Comparing common INR and VAE based learned models for the hologram compression task in a split view. For every image inset, the lower triangles represent the compressed phase-only hologram with PSNR and SSIM metrics, respectively. The upper triangle displays the 3D reconstruction of the corresponding 3D phase-only hologram, incorporating PSNR, SSIM ,and LPIPS metrics in order (Source image: photosteve101).

Abstract

We evaluate the performance of four common learned models utilizing INR and VAE structures for compressing phase-only holograms in holographic displays. The evaluated models include a **vanilla MLP**, **SIREN** [Sitzmann et al. 2020], and **FilmSIREN** [Chan et al. 2021], with **TAESD** [Bohan 2023] as the representative VAE model. Our experiments reveal that a pretrained image VAE, **TAESD**, with $2.2M$ parameters struggles with phase-only hologram compression, revealing the need for task-specific adaptations. Among the INRs, **SIREN** with $4.9k$ parameters achieves %40 compression with high quality in the reconstructed 3D images (PSNR = 34.54 dB). These results emphasize the effectiveness of INRs and identify the limitations of pretrained image compression VAEs for hologram compression task.

CCS Concepts

• **Computing methodologies** → **Image compression**; • **Theory of computation** → **Data compression**; • **Hardware** → *Displays and imagers*; *Emerging optical and photonic technologies*.

Keywords

Computer-Generated Holography, Hologram Compression, Holographic Displays

1 Introduction

Emerging holographic displays [Kavaklı et al. 2023; Zheng et al. 2024] utilize phase-only holograms to reconstruct full-color 3D scenes at various optical depths. The diversity of phase-only hologram types and encoding schemes available today reflects the infancy of display hardware and the lack of standardized practices. These hologram types, which include single-color and multi-color variants, and their counterpart encoding schemes like double-phase or direct phase encoding, require efficient compression due to their high-frequency information content (see Fig. 2). On the other hand, learned models [Wang et al. 2022] are known to struggle preserving high-frequency features, and the potential benefits of compression using learned models remain uncertain, regardless of type or encoding. This ambiguity motivates our investigation into whether learned models can effectively compress phase-only holograms and contribute to improved storage and transmission efficiency. Therefore, we chose two general learned
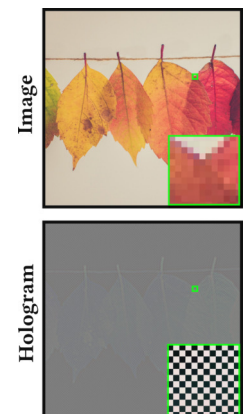

Figure 2: Unlike natural images, holograms are dominated by high-frequency content. (Source image: leaves-Color2025)

model structures: VAE, which generates images by learning in a latent space, and INR, which directly represents image content using implicit neural functions. Our work presents the first attempt to assess the learned models in a structured way for the hologram compression task. We focus solely on comparing the performance of different existing methods for hologram image compression with the aim of providing a foundation for future optimization using novel models or approaches. Specifically, in this early study, we analyze a **vanilla MLP**, **SIREN** [Sitzmann et al. 2020], and **FilmSIREN** [Chan et al. 2021], with **TAESD** [Bohan 2023] as the representative VAE model.

2 Method

Our assessments involve using single-color double-phase encoded phase-only holograms, $P \in \mathbb{R}^{3 \times 512 \times 512}$, using three color primaries. These Ps are calculated for three wavelengths, $\lambda = \{473, 515, 639\}\,nm$ and a fixed pixel pitch, $px = 3.74\,\mu m$ (Jasper Display JD7714). We adopt an off-the-shelf **TAESD** trained for image compression task. Specifically, the **TAESD** with $2.2M$ parameters encodes P to a $\alpha \in \mathbb{R}^{16 \times 64 \times 64}$ and later decodes into the original resolution of $3 \times 512 \times 512$. Fig. 1 shows pretrained **TAESD** fails, requiring dedicated training for generalization, the compression would amount to %92 reduction (excluding **TAESD** params). Thus, we choose to explore INR based models to see if the feature size could be further reduced while accepting longer training times as INRs typically are overfitted on a single data at a time. In our study, we compare three foundational INR architectures (**vanilla MLP**, **SIREN**, and **FilmSIREN**), aiming for ∼ %40 compression to strike a balance between the quality of the reconstructed image and the compression ratio. Ps are split into patches (e.g., $3 \times 64 \times 64$), a separate model is trained for each patch (initialized from prior weights), and their outputs are combined for full reconstruction. Experiments that we are going to detail in the next section utilize ten different holograms (The purpose of selecting a small but diverse set of initial samples in this study is to demonstrate the comparative trends among different methods. The large-scale validation work will be addressed in the subsequent research.) and turns them into patches by following the choices listed in Tbl. 1. All INRs use Adam (lr=0.0001) with StepLR (gamma=0.5 every 5000 epochs), trained for 10000 epochs. More details of our models are available in our supplementary.

3 Results and Discussions

SIREN and **FilmSIREN** provide strong compression, outperforming **vanilla MLP**, with **SIREN** showing best consistency. In our current experiments in Tbl. 1, **SIREN** achieves the highest fidelity with a PSNR of 42.29 dB and SSIM of 0.9997 at $3 \times 64 \times 64$ patch size, For 3D holographic reconstructions with a 5 mm volume depth, metrics are averaged over three focal planes at propagation distances of -2.5, 0, and $+2.5$ mm. Under this setting, **SIREN** attains a PSNR of 34.54 dB, SSIM of 0.96, LPIPS of 0.10 marginally outperforms **FilmSIREN** (PSNR = 33.27 dB, SSIM = 0.94, LPIPS = 0.15). Additionally, under identical training schedules, both **SIREN** and **FilmSIREN** frequently satisfied the early stopping criterion near 2000 epochs. This consistency implies a relatively smooth optimization process, suggesting that these models can converge effectively

without compromising image quality, which is a favorable property in hologram compression task. The computational demand of around 40 minutes per hologram seems justified by **SIREN**'s ability to preserve quality. These observations suggest that specialized INR architectures require further investigation for the hologram compression task, potentially opening new solutions for efficient 3D scene representation in holographic displays. Achieving robust compression remains an open challenge; our study guides future work on efficient 3D holographic rendering/storage.

Table 1: Patch based hologram quality comparison between vanilla MLP, FilmSIREN, and SIREN.

vanilla MLP			
Patch size	PSNR ± Std.	Params	Comp. Ratio
$3 \times 64 \times 64$	40.06 ± 2.73	5,059	41%
$\mathbf{3 \times 96 \times 96}$	**41.50 ± 2.91**	**11,139**	**40%**
$3 \times 128 \times 128$	39.88 ± 2.05	19,459	40%
$3 \times 160 \times 160$	40.71 ± 1.89	31,939	41%
FilmSIREN			
Patch size	PSNR ± Std.	Params	Comp. Ratio
$\mathbf{3 \times 64 \times 64}$	**40.92 ± 2.91**	**4,869**	**40%**
$3 \times 96 \times 96$	40.68 ± 2.58	10,755	39%
$3 \times 128 \times 128$	39.70 ± 3.18	19,137	39%
$3 \times 160 \times 160$	35.48 ± 2.93	30,357	40%
SIREN			
Patch size	PSNR ± Std.	Params	Comp. Ratio
$\mathbf{3 \times 64 \times 64}$	**42.29 ± 2.45**	**4,899**	**40%**
$3 \times 96 \times 96$	40.83 ± 2.63	11,171	40%
$3 \times 128 \times 128$	39.32 ± 3.08	19,491	40%
$3 \times 160 \times 160$	37.51 ± 4.88	31,971	41%

References

Ollin Boer Bohan. 2023. Tiny Autoencoder for Stable Diffusion. https://github.com/madebyollin/taesd. Accessed: June 2025.

Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. 2021. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5799–5809. doi:10.1109/CVPR46437.2021.00574

Koray Kavaklı, Liang Shi, Hakan Urey, Wojciech Matusik, and Kaan Akşit. 2023. Multi-color Holograms Improve Brightness in Holographic Displays. In *SIGGRAPH ASIA 2023 Conference Papers* (Sydney, NSW, Australia) *(SA '23)*. Article 20, 11 pages. doi:10.1145/3610548.3618135

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems* 33 (2020), 7462–7473. doi:10.48550/arXiv.2006.09661

Yujie Wang, Praneeth Chakravarthula, Qi Sun, and Baoquan Chen. 2022. Joint neural phase retrieval and compression for energy-and computation-efficient holography on the edge. *ACM Transactions on Graphics* 41, 4 (2022). doi:10.1145/3528223.3530070

Chuanjun Zheng, Yicheng Zhan, Liang Shi, Ozan Cakmakci, and Kaan Akşit. 2024. Focal Surface Holographic Light Transport using Learned Spatially Adaptive Convolutions. In *SIGGRAPH Asia 2024 Technical Communications (SA Technical Communications '24)* (Tokyo, Japan) *(SA '24)*. doi:10.1145/3681758.3697989