Deep-Learning-based Phase Retrieval with Time-Multiplexing for Holographic Display

Weiwei Wu* (wwwu@stanford.edu)

Introduction

Computer-generated holography (CGH) aims to synthesize phase-only holograms that reconstruct desired optical fields for 3D display. Recent Gaussian Wave Splatting (GWS) methods connect Gaussian-based scene representations to hologram formation, enabling efficient primitive-based CGH. However, achieving both high spatial image quality and view-dependent angular effects remains challenging, especially when current random-phase time-multiplexing methods require many frames to converge.

Our idea: formulate this as a phase-retrieval problem with joint spatial and angular constraints. We predict a strong single-frame phase from (S, V), where S is the spatial intensity target and V is the Fourier / angular intensity target, then apply lightweight time-multiplexing refinement.

Motivation: Spatial vs Angular Control in Gaussian Wave Splatting

Gaussian Wave Splatting (GWS) converts Gaussian primitives into holograms by matching two physical constraints:

- a desired **spatial intensity footprint** on the SLM plane $S(\mathbf{x})$
- a desired **angular / Fourier emission profile** $V(\mathbf{k})$

Spatial footprint (Eq. S24) [1].

For a single Gaussian primitive, the SLM-plane spectrum can be written as

$$\hat{u}(\mathbf{k}) = \det(\mathbf{J})\det(\mathbf{S})\hat{G}(\mathbf{S}\mathbf{R}^{-1}\mathbf{k})\,e^{j\mathbf{k}\cdot\boldsymbol{\mu}}$$

which defines the spatial Gaussian intensity pattern

$$S(\mathbf{x}) = \left|\mathcal{F}^{-1}\{\hat{u}(\mathbf{k})\}\right|^2.$$

This formulation preserves Gaussian spatial structure but provides **no control over the angular spectrum**.

Angular control via random phase (Eq. S40) [1,2].

Random-phase Gaussian Wave Splatting (RP-GWS) introduces view-dependent effects by convolving the primitive spectrum with an angular emission kernel $V(\mathbf{k})$ and random phase:

$$\hat{u}_{RP}(\mathbf{k}) = \hat{u}(\mathbf{k}) * \left(V(\mathbf{k})e^{j\phi^{(t)}(\mathbf{k})}\right)$$

Multiple hologram frames are averaged through time multiplexing (TM):

$$S(\mathbf{x}) \approx \frac{1}{T}\sum_{t=1}^{T}\left|\mathcal{F}^{-1}\{\hat{u}^{(t)}(\mathbf{k})\}\right|^2$$

to recover the desired spatial intensity.

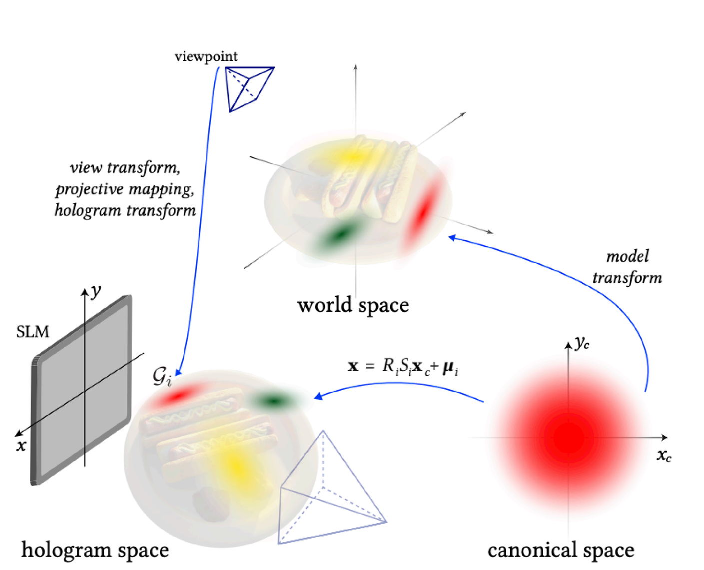

Fig. 1a. Gaussian primitive mapping from world space to hologram space in Gaussian Wave Splatting [1].

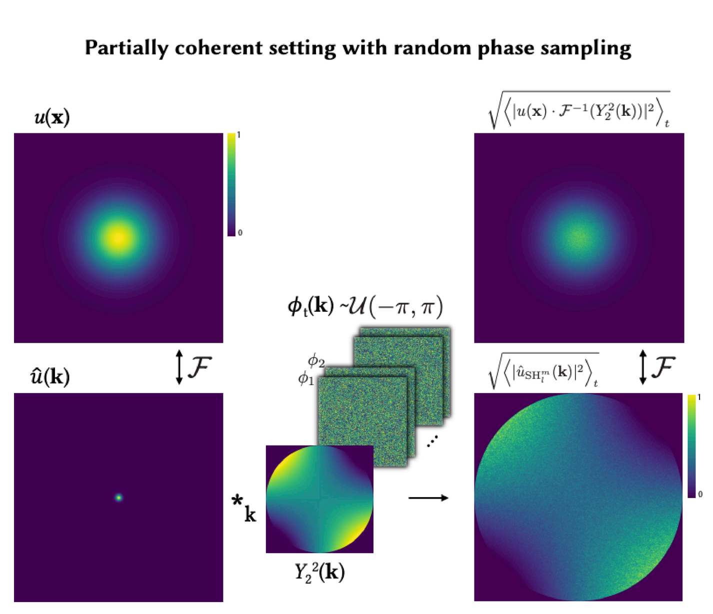

Fig. 1b. Random-phase Gaussian Wave Splatting introduces angular emission control using random phase and time multiplexing [1].

Limitation.

Finite-frame time multiplexing introduces residual error in both the spatial intensity $S(\mathbf{x})$ and the angular spectrum $V(\mathbf{k})$, often requiring many frames to converge.

Can we achieve higher spatial fidelity and view-dependent effects with fewer frames?

New Technique: Learned Phase Retrieval with Lightweight TM Refinement

Phase-retrieval formulation. Instead of satisfying the spatial and angular targets only through random-phase averaging, we explicitly enforce the Fourier / angular target and solve only for the phase:

$$\hat{u}(\mathbf{k}) = \sqrt{V(\mathbf{k})}\,e^{i\phi(\mathbf{k})}, \qquad |\hat{u}(\mathbf{k})|^2 = V(\mathbf{k}).$$

Therefore, the angular spectrum is satisfied **exactly by construction**, and the remaining problem is to find a phase $\phi(\mathbf{k})$ such that

$$S(\mathbf{x}) = \left|\mathcal{F}^{-1}\{\hat{u}(\mathbf{k})\}\right|^2.$$

Frame 1: neural phase prediction. Given the spatial target $S(\mathbf{x})$ and angular target $V(\mathbf{k})$, a neural network predicts a strong initial phase

$$\phi_0(\mathbf{k}) = f_\theta(S(\mathbf{x}), V(\mathbf{k}))\,.$$

The first-frame spectrum is then constructed as

$$\hat{u}^{(1)}(\mathbf{k}) = \sqrt{V(\mathbf{k})}\,e^{i\phi_0(\mathbf{k})}.$$

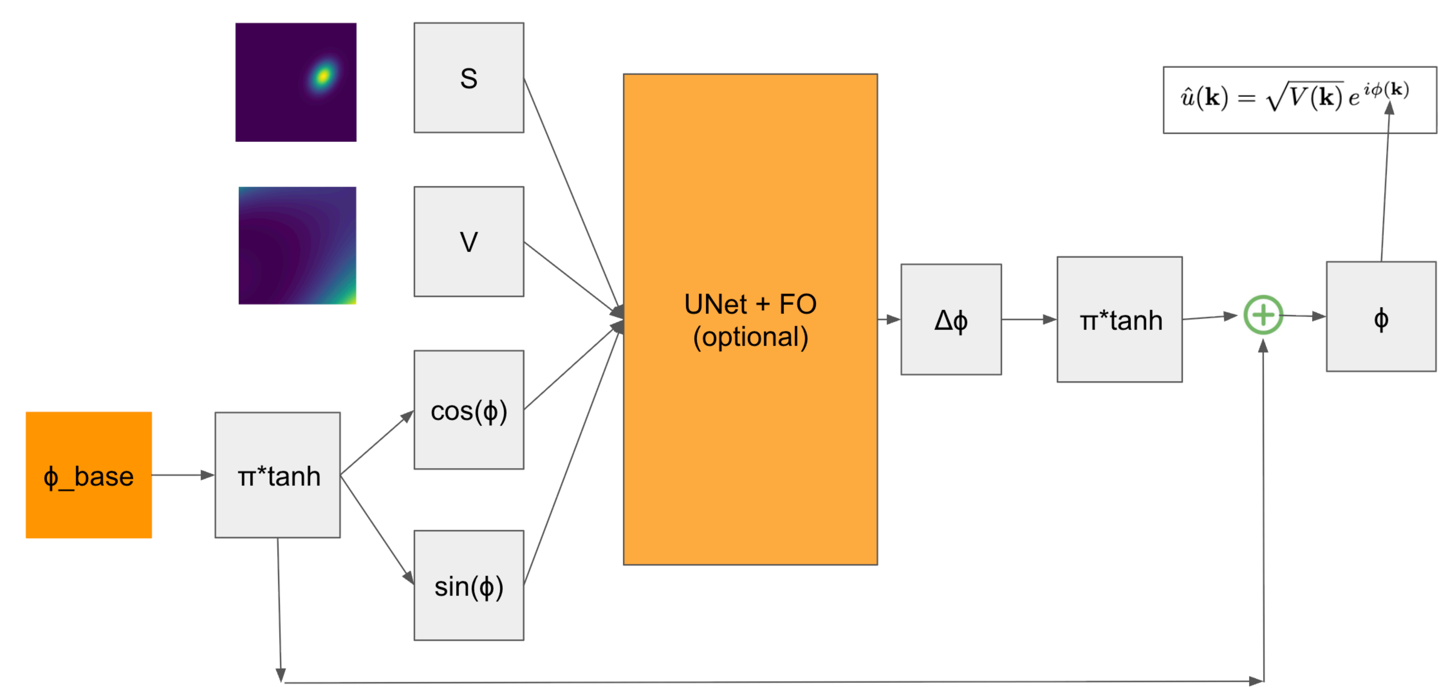

Fig. 3. Model structure for generating the first-frame phase from (S, V) and a learnable base phase prior. Orange sections, the model and a phase base, are trainable

Frames 2–N: lightweight TM refinement. Additional frames are generated without further neural inference. Starting from the learned phase ϕ_0, we apply a small phase perturbation, one spatial projection step, and then re-impose the exact Fourier magnitude $\sqrt{V(\mathbf{k})}$.

$$\begin{aligned}
\tilde{\phi}^{(t)}(\mathbf{k}) &= \mathrm{wrap}\Big(\phi_0(\mathbf{k}) + \epsilon_\phi\eta^{(t)}(\mathbf{k})\Big)\,,\\
\hat{u}^{(t)}(\mathbf{k}) &= \sqrt{V(\mathbf{k})}\,e^{i\tilde{\phi}^{(t)}(\mathbf{k})},\\
u^{(t)}(\mathbf{x}) &= \mathcal{F}^{-1}\Big\{\hat{u}^{(t)}(\mathbf{k})\Big\}\,,\\
u^{(t)}_{\mathrm{proj}}(\mathbf{x}) &= \sqrt{S^{(t)}_{\mathrm{tgt}}(\mathbf{x})}\,e^{i\angle u^{(t)}(\mathbf{x})},\\
\hat{u}^{(t)}_{\mathrm{proj}}(\mathbf{k}) &= \mathcal{F}\Big\{u^{(t)}_{\mathrm{proj}}(\mathbf{x})\Big\}\,,\\
\phi^{(t)}(\mathbf{k}) &= \angle\hat{u}^{(t)}_{\mathrm{proj}}(\mathbf{k}),\\
\hat{u}^{(t)}(\mathbf{k}) &= \sqrt{V(\mathbf{k})}\,e^{i\phi^{(t)}(\mathbf{k})}.
\end{aligned}$$

where

$$\begin{aligned}
\bar{S}^{(t)}(\mathbf{x}) &= \frac{1}{t}\sum_{\tau=1}^{t}\left|u^{(\tau)}(\mathbf{x})\right|^2,\\
R^{(t)}(\mathbf{x}) &= S(\mathbf{x}) - \bar{S}^{(t)}(\mathbf{x}),\\
S^{(t+1)}_{\mathrm{tgt}}(\mathbf{x}) &= \Pi_{\geq 0}\Big(S(\mathbf{x}) + \gamma R^{(t)}(\mathbf{x})\Big)\,.
\end{aligned}$$

Relatively cheap (WGS-like)

The final reconstruction averages the T frame intensities:

$$\hat{S}(\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T}|u^{(t)}(\mathbf{x})|^2.$$

High-quality first-frame prediction + cheap TM refinement = fewer frames, exact $V(\mathbf{k})$, and improved spatial fidelity.

Experimental Results

- **1 frame OURS** achieves spatial quality comparable to roughly **15–20 frames RP-GWS**, while satisfying the angular target $V(\mathbf{k})$ exactly.
- At **20 frames**, OURS achieves nearly an **order-of-magnitude lower dev-set MSE** than naive TM.
- Inference is in the **millisecond regime per sample**; real deployment can use larger batch size for higher throughput.

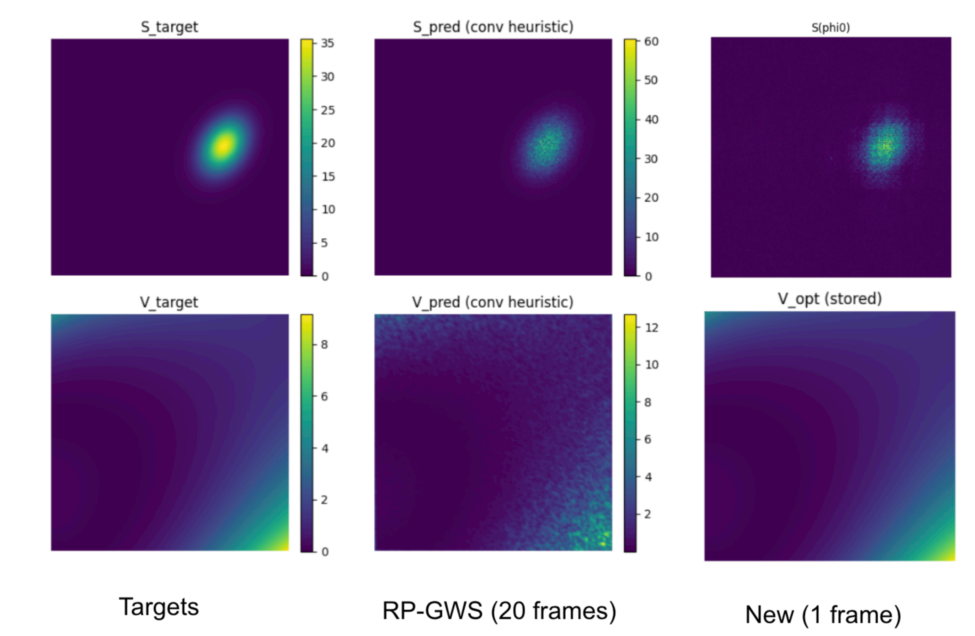

Fig. 4. One learned frame already matches the spatial quality of roughly 15–20 RP-GWS frames, while preserving the exact angular target.

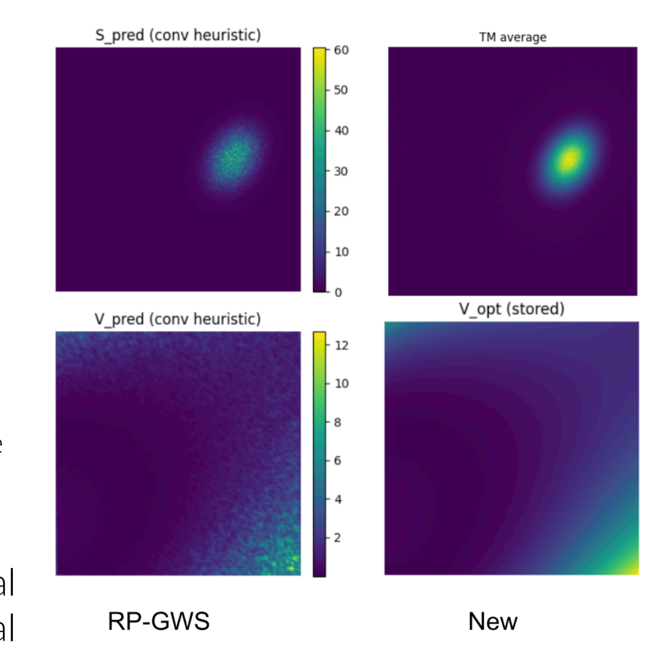

Fig. 5. Comparison at 20 frames.

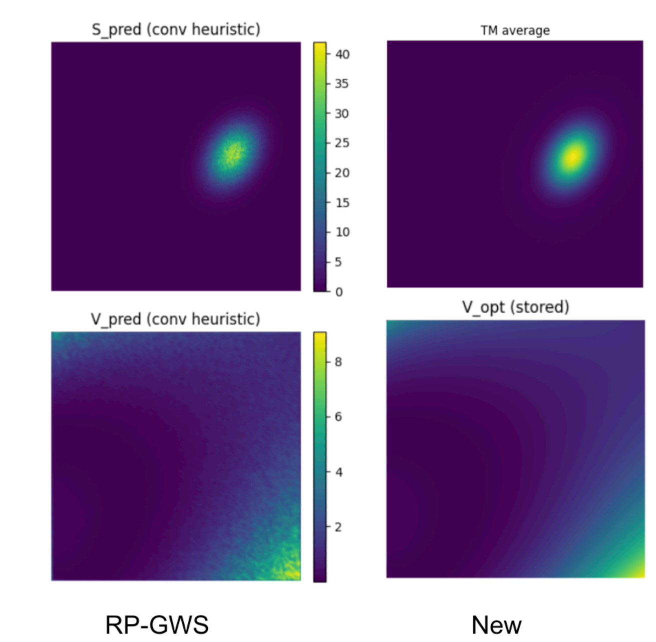

Fig. 6. Comparison at 128 frames.

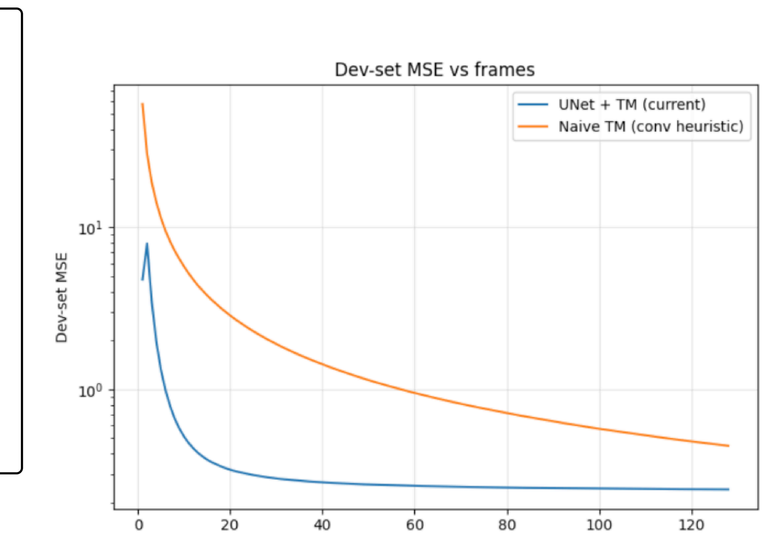

Fig. 7. Dev-set MSE vs. frame count. OURS converges much faster and reaches lower error.

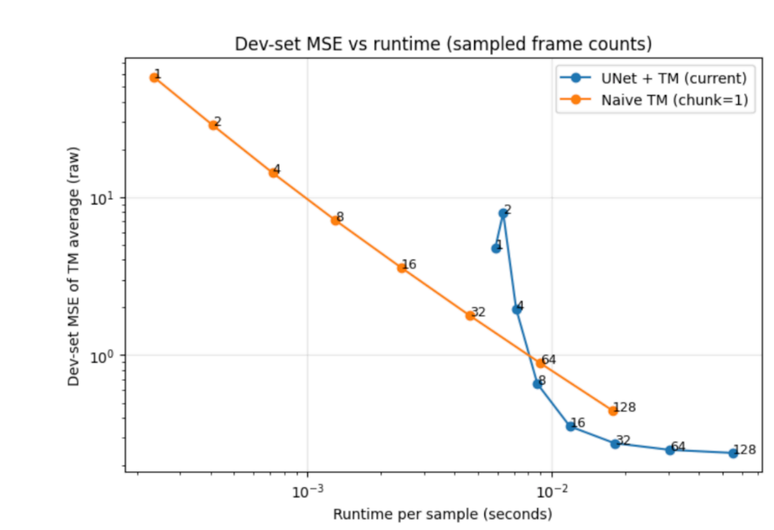

Fig. 8. Dev-set MSE vs. runtime per sample (batch = 1). OURS operates in the millisecond regime.

Conclusion & Outlook

We reformulate Gaussian Wave Splatting hologram synthesis as a phase-retrieval problem that enforces the angular spectrum $V(\mathbf{k})$ exactly while learning a high-quality single-frame phase. With lightweight TM refinement, one learned frame achieves spatial quality comparable to 15–20 RP-GWS frames.

Future work will extend this formulation to full scene-level hologram rendering (around 500k Gaussian primitives per scene).

References

[1] S. Choi, B. Chao, J. Yang, M. Gopakumar, G. Wetzstein. *Gaussian Wave Splatting for Computer-Generated Holography.* ACM Transactions on Graphics (SIGGRAPH), 2025.

[2] B. Chao, J. Yang, S. Choi, M. Gopakumar, R. Koiso, G. Wetzstein. *Random-phase Gaussian Wave Splatting for Computer-Generated Holography.* 2025.