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

Weiwei Wu

1 Motivation

Gaussian Wave Splatting (GWS) converts a 2D Gaussian-primitive scene into a hologram by matching, for each primitive, a desired spatial intensity footprint on the SLM plane and a desired angular / Fourier emission profile. Concretely, for a single Gaussian primitive, the target spatial profile is the squared-magnitude of an inverse Fourier transform:

$$S(x, y) \ = \ \left|\mathcal{F}^{-1}\{\hat{u}(k_x, k_y)\}\right|^2,$$

and the target angular/Fourier profile is encoded as a spherical-harmonics angular spectrum (often denoted as the SH "kernel" / "Q"):

$$V(k_x, k_y) \ \equiv \ Q(k_x, k_y) \quad \text{(desired Fourier magnitude / emission profile).}$$

Smooth-phase GWS achieves high in-focus quality but tends to under-utilize bandwidth, limiting parallax and natural defocus. Random-phase GWS (RP-GWS) fixes this and addes view-dependent effects by convolving the primitive spectrum with an angular-emission kernel and using time multiplexing (TM) with random phase to suppress speckle:

$$\hat{u}_{\mathrm{RP}}(k) \ = \ \hat{u}(k) \ * \ \left(Q(k)\, e^{j\phi^{(t)}(k)}\right), \qquad t = 1, \ldots, T,$$

so that spatial-domain Gaussian structure is preserved (in expectation / over TM) while achieving the desired Fourier-domain emission profile $Q(k)$.

What we will do. We will replace the convolution formulation with random phase with a deep phase-retrieval model that predicts a high-quality initial phase φ under joint constraints from (S, V), then apply a lightweight TM refinement pipeline to reach high image quality with fewer iterations/frames than the TM method proposed in the RP-GWS paper.

2 Related work

- Suyeon Choi, Brian Chao, Jacqueline Yang, Manu Gopakumar, Gordon Wetzstein. *Gaussian Wave Splatting for Computer-Generated Holography.* ACM TOG / SIGGRAPH 2025. :contentReference[oaicite:2]index=2
- Brian Chao, Jacqueline Yang, Suyeon Choi, Manu Gopakumar, Ryota Koiso, Gordon Wetzstein. *Random-phase Gaussian Wave Splatting for Computer-Generated Holography.* (RP-GWS / GWS-RP). :contentReference[oaicite:3]index=3

3 Method: Learned single-frame phase from (S, V) with residual-steered TM refinement

3.1 Overview and key idea

Existing RP-GWS-style rendering achieves spatial fidelity mainly *in expectation* over time-multiplexed frames, so finite-T results can retain speckle and converge slowly. Our method shifts the burden of quality to a strong *single-frame* learned phase, then uses a cheap TM refinement to remove residual bias while preserving the angular/Fourier target exactly at every frame.

3.2 Exact enforcement of the angular/Fourier target

At render time we explicitly construct the angular spectrum as

$$\hat{u}(\boldsymbol{k}_\perp) = \sqrt{V(\boldsymbol{k}_\perp)}\, e^{\mathrm{i}\phi(\boldsymbol{k}_\perp)}, \tag{1}$$

so that $|\hat{u}(\boldsymbol{k}_\perp)|^2 = V(\boldsymbol{k}_\perp)$ holds by construction. All remaining error reduction is performed via optimizing the phase ϕ.

3.3 Single-frame learned phase synthesis (Frame 1)

Given a spatial target $S(\boldsymbol{x})$ and Fourier-domain target $V(\boldsymbol{k}_\perp)$, a neural network predicts an initial phase:

$$\phi_0(\boldsymbol{k}_\perp) = f_\theta(S(\boldsymbol{x}), V(\boldsymbol{k}_\perp))\,. \tag{2}$$

We form the first-frame spectrum and reconstruction:

$$\hat{u}^{(1)}(\boldsymbol{k}_\perp) = \sqrt{V(\boldsymbol{k}_\perp)}\, e^{\mathrm{i}\phi_0(\boldsymbol{k}_\perp)}, \tag{3}$$

$$u^{(1)}(\boldsymbol{x}) = \mathcal{F}^{-1}\{\hat{u}^{(1)}\}(\boldsymbol{x}), \qquad S^{(1)}(\boldsymbol{x}) = \left|u^{(1)}(\boldsymbol{x})\right|^2. \tag{4}$$

Training minimizes a spatial-domain loss (e.g., $\|S^{(1)} - S\|$ plus regularizers), yielding a high-quality single-frame initialization that already satisfies V exactly.

3.4 Cheap TM refinement (Frames $t \geq 2$) with one-step projection

For additional TM frames we generate phases without further neural inference. For each frame we start from a perturbed phase and apply a single alternating-projection step to reduce spatial error:

$$\tilde{\phi}^{(t)}(\boldsymbol{k}_\perp) = \mathrm{wrap}\Big(\phi_0(\boldsymbol{k}_\perp) + \epsilon_\phi\, \eta^{(t)}(\boldsymbol{k}_\perp)\Big)\,, \tag{5}$$

$$\hat{u}^{(t)}(\boldsymbol{k}_\perp) = \sqrt{V(\boldsymbol{k}_\perp)}\, e^{\mathrm{i}\tilde{\phi}^{(t)}(\boldsymbol{k}_\perp)}, \tag{6}$$

$$u^{(t)}(\boldsymbol{x}) = \mathcal{F}^{-1}\{\hat{u}^{(t)}\}(\boldsymbol{x}), \tag{7}$$

$$u^{(t)}_{\mathrm{proj}}(\boldsymbol{x}) = \sqrt{S^{(t)}_{\mathrm{tgt}}(\boldsymbol{x})}\, e^{\mathrm{i}\angle u^{(t)}(\boldsymbol{x})}, \tag{8}$$

$$\hat{u}^{(t)}_{\mathrm{proj}}(\boldsymbol{k}_\perp) = \mathcal{F}\{u^{(t)}_{\mathrm{proj}}\}(\boldsymbol{k}_\perp), \tag{9}$$

$$\phi^{(t)}(\boldsymbol{k}_\perp) = \angle \hat{u}^{(t)}_{\mathrm{proj}}(\boldsymbol{k}_\perp), \tag{10}$$

$$\hat{u}^{(t)}(\boldsymbol{k}_\perp) = \sqrt{V(\boldsymbol{k}_\perp)}\, e^{\mathrm{i}\phi^{(t)}(\boldsymbol{k}_\perp)}. \tag{11}$$

Because the final step explicitly re-imposes the Fourier magnitude $\sqrt{V}$, V **remains exact for every frame**.

3.5 Residual-steered spatial targets to correct mean bias

Averaging across TM frames suppresses random speckle, but correlated bias can persist. We therefore steer the per-frame spatial target using the running-average residual:

$$\bar{S}^{(t)}(\boldsymbol{x}) = \frac{1}{t}\sum_{\tau=1}^{t}\left|u^{(\tau)}(\boldsymbol{x})\right|^2, \tag{12}$$

$$R^{(t)}(\boldsymbol{x}) = S(\boldsymbol{x}) - \bar{S}^{(t)}(\boldsymbol{x}), \tag{13}$$

$$S_{\text{tgt}}^{(t+1)}(\boldsymbol{x}) = \Pi_{\geq 0}\Big(S(\boldsymbol{x}) + \gamma\, R^{(t)}(\boldsymbol{x})\Big), \tag{14}$$

where $\Pi_{\geq 0}$ clips to nonnegative intensities. In practice, we optionally apply ROI masking, mild smoothing, and energy renormalization to stabilize updates. This residual-steered schedule directly targets the *mean* error, avoiding early plateaus and improving convergence for small T.

3.6 Output

The final reconstruction is the average of T intensity frames:

$$\hat{S}^{(T)}(\boldsymbol{x}) = \frac{1}{T}\sum_{t=1}^{T}\left|u^{(t)}(\boldsymbol{x})\right|^2, \tag{15}$$

with **exact** per-frame angular constraint $\left|\hat{u}^{(t)}\right|^2 = V$ and progressively improved spatial match driven by the learned initializer and residual-steered TM refinement.