Transformer Enabled Dual-Comb Ghost Imaging for Optical Fiber Sensing

David Dang[1,3] **Myoung-Gyun Suh**[2] **Maodong Gao**[2] **Dong-Chel Shin**[2]
Aggraj Gupta[2] **Byoung Jun Park**[2] **Can Uzundal**[1] **Beyonce Hu**[1]
Yucheng Jin[1] **Wilton J.M. Kort-Kamp**[3] **Ho Wai (Howard) Lee**[1]
[1]University of California, Irvine [2]NTT Physics and Informatics Laboratories
[3]Los Alamos National Laboratory
{dangd5,beyonceh,yuchej9,Howardhw.lee}@uci.edu
{myoung-gyun.suh,byoungjun.park,can.uzundal}@ntt-research.com,
mgao@caltech.edu, dongchel@mit.edu, kortkamp@lanl.gov

Abstract

Ghost imaging (GI) reconstructs images from single-pixel measurements but remains hindered by slow pattern projection and noise-sensitive reconstruction. We present a dual-comb ghost imaging framework that addresses these limitations. Dual optical frequency combs generate hundreds of uncorrelated speckle patterns in parallel, enabling snapshot bucket-sum detection through a single-core fiber without spatial or spectral scanning. For recovery, we introduce Optical Ghost-GPT, a transformer-based model that achieves real-time, high-fidelity reconstruction at ultra-low sampling ratios. This combination of dual-comb hardware and deep learning significantly outperforms classical GI in speed, robustness, and image quality. As a proof of concept, we highlight fiber-based endoscopy as a key application, where our approach could deliver minimally invasive, high-resolution imaging with sub-millimeter hardware.

1 Introduction

Ghost imaging circumvents the need for conventional 2-D array sensor cameras by replacing them with correlated intensity measurements from a single-pixel (bucket) detector [1]. In a typical setup, a sequence of known illumination patterns is projected onto the sample, and the total transmitted or reflected intensity is measured. Statistical correlations between the projected patterns and the measured signals are then used to computationally reconstruct the image. This framework has demonstrated key advantages: enhanced signal-to-noise ratios in sparse light conditions [2, 3, 4], robustness to scattering and turbid media [5, 6], and inherent compatibility with single-pixel detectors. These properties make ghost imaging particularly attractive for fiber-based endoscopy, where conventional 2D detector arrays are difficult to integrate, and miniaturization is critical for patient comfort and access to small or delicate anatomical regions [7, 8].

Despite these advantages, classical ghost imaging suffers from fundamental bottlenecks in acquisition and reconstruction. Each illumination pattern is typically projected sequentially, requiring thousands of measurements to achieve acceptable fidelity. Reconstruction methods, often based on iterative correlation or compressive sensing, converge slowly, degrade in the presence of noise, and fail at low sampling ratios—making real-time, high-resolution imaging impractical for clinical use.

To overcome these challenges, we propose a dual-pronged solution. First, we integrate dual optical frequency comb (dual-comb) interferometry [9, 10, 11] with compressive ghost imaging to generate snapshot speckle illumination and detection through a single-core fiber, thereby eliminating the need for slow spatial or spectral scanning. Second, we introduce Optical Ghost-GPT, a transformer-based

reconstruction framework that leverages global attention mechanisms to rapidly and accurately map illumination–signal correlations. This enables real-time, high-fidelity imaging even under noisy and undersampled conditions.

Our contributions include: (1) the first demonstration of dual-comb–enabled ghost imaging in a fiber-compatible architecture with deep learning–based reconstruction, (2) significant improvements in speed (greater than 60 FPS) at ultra-low sampling ratios (down to 0.29%) . Together, these advances establish ghost imaging as a promising strategy for next-generation fiber-based endoscopy, offering high-performance imaging in minimally invasive and resource-constrained settings.

2 Dual-Comb Ghost Imaging: Experimental Details

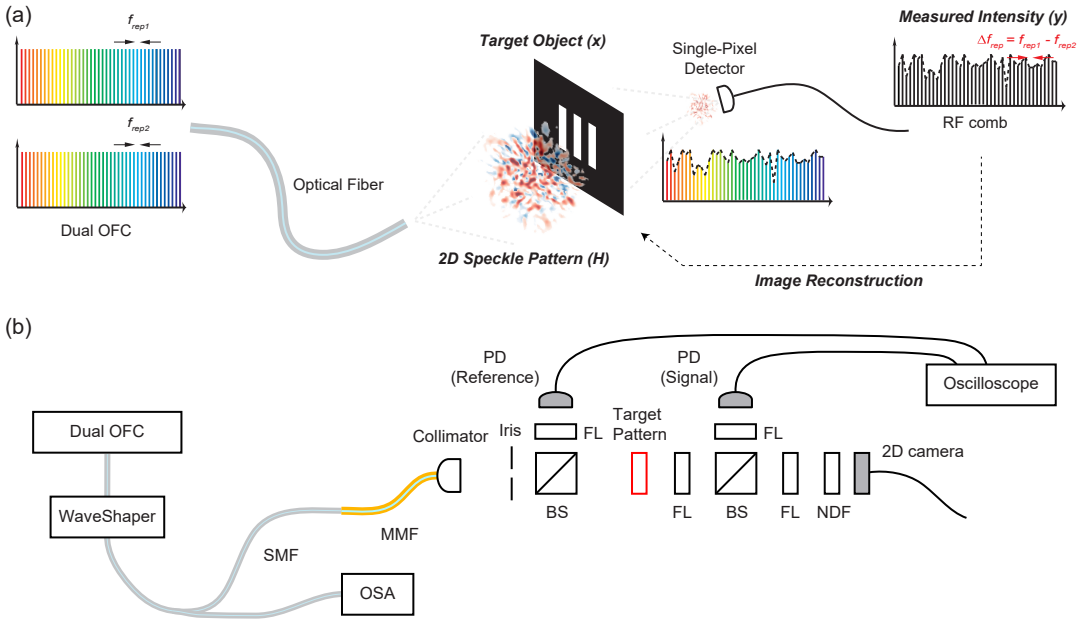

Figure 1: **Optical Fiber-based Dual-Comb Ghost Imaging.** (a) Conceptual illustration of optical fiber-based ghost imaging using dual optical frequency combs. Each comb line generates an uncorrelated speckle pattern at the fiber tip, and the set of speckle patterns is mapped onto the target object x. The encoded speckle pattern ($H \times x$) is collected by a single-pixel detector, and the resulting dual-comb interference signal ("interferogram") is recorded. (b) Experimental setup diagram. SMF: Single Mode Fiber, OSA: Optical Spectrum Analyzer, PD: Photodetector, FL: Focusing Lens, BS: Beam Splitter, NDF: Neutral Density Filter.

An optical frequency comb (OFC) consists of equally spaced, mutually coherent frequency lines. In our approach, dual OFCs generate a set of uncorrelated 2D speckle patterns H at the fiber tip, which are projected onto the target object x. The encoded intensity ($H \times x$) is measured by a single-pixel detector, producing a dual-comb interferogram. A fast Fourier transform (FFT) converts this time-domain signal into frequency-domain bucket-sum data $y = H \times x$, enabling parallelized speckle projection and significantly increased imaging speed.

Two electro-optic (EO) combs with slightly different free spectral ranges (f_{FSR} and $f_{\mathrm{FSR}} + \Delta f_{\mathrm{FSR}}$) are generated from a CW laser at $1550\,\mathrm{nm}$. After frequency shifting with acousto-optic modulators, the EO modulators produce combs at $f_{\mathrm{FSR}} = 20\,\mathrm{GHz}$ with $\Delta f_{\mathrm{FSR}} = 200\,\mathrm{Hz}$. Recombination yields RF-domain beat signals upon photodetection, allowing fast and precise electronic measurement of the bucket intensities as well as with coherent averaging for improved SNR. A Waveshaper enables selection of individual comb lines to allow for the measurement of the speckle patterns in the inital calibration procedure. Light is coupled into a $200\,\mu$m-core multimode fiber, producing hundreds of modes and corresponding speckle patterns. These are projected through the target, split into signal

and reference arms, and detected with InGaAs photodiodes (500 kHz bandwidth). Speckle patterns are calibrated separately with a 256×256 InGaAs camera.

Approximately 200 comb lines spanning 1530–1565 nm were used, yielding speckle patterns that were uncorrelated across wavelengths. A negative USAF 1951 resolution chart served as the target. To overcome the slow image reconstruction speed on the computational side we, we developed **Optical Ghost-GPT**, a transformer-based reconstruction model that strongly suppresses noise and enhances image fidelity.

3 Transformer Modeling

Transformers, originally developed for natural language processing [12], have since become a foundation for computer vision [13], speech [14], and scientific data analysis [15]. Their self-attention mechanism models long-range dependencies while enabling efficient parallel computation, powering state-of-the-art models such as BERT, GPT [16], and Vision Transformers (ViTs) [17, 18]. In ViTs, images are divided into patches that act as tokens. To adapt this idea to dual-comb ghost imaging, we instead define each token by concatenating a flattened illumination pattern from the sensing matrix Ψ with its corresponding bucket value.

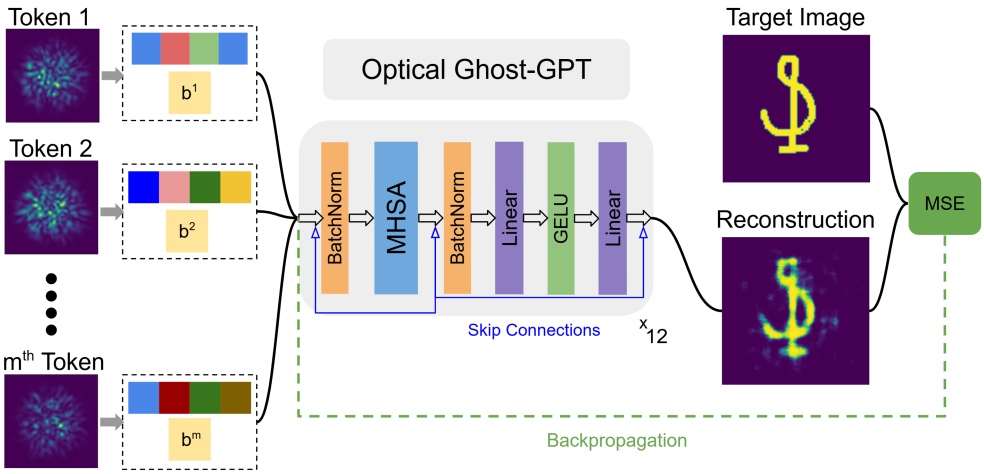

Figure 2: **Schematic of Optical Ghost-GPT**. The speckle patterns are first compressed into a latent space and then concatenated with the corresponding bucket measurements to form the input token sequence.

Model Architecture. Each token consists of a speckle pattern (flattened to 256×256) linearly projected to a smaller embedding dimensions concatenated with its bucket sum, and augmented with learned positional embeddings (context size $C = 250$). A stack of $L = 12$ transformer blocks with multi-head self-attention ($H = 32$) and feedforward layers (GELU activations) processes the sequence, with dropout (0.1), residual connections, and batch normalization. Final token representations are projected to $\mathbb{R}^{16}$, flattened, and mapped back to the original 256×256 image with a sigmoid output.

Training. Calibration speckle patterns (188 total) are collected with a 2D camera and convolved with MNIST/Omniglot images to generate synthetic bucket sums. The network is trained to reconstruct the target image from speckle–bucket pairs using mean squared error (MSE) loss and AdamW optimizer.

4 Ghost-GPT Reconstruction Results

To test our trained transformer model, we performed reconstructions on experimental optics data taken with our setup. Figure 3 and Table 1 shows the reconstruction results of a stripe (top) and the number 2 (bottom) compared with classical algorithms and Ghost-GPT and the corresponding

MSE and SSIM values. In real world environments, we demonstrate that Ghost-GPT is able to successfully reconstruct images with higher fidelity for both targets compared to classical methods, while reconstructing images in approximately 8 ms (order of magnitudes faster than the iterative classical algorithms). In addition to comparisons with classical algorithm, we also explored other deep learning models, such as CNN and U-Net architectures, and found that our transformer model outperformed by 46.2 % and 36.4% respectively, based on mean squared error.

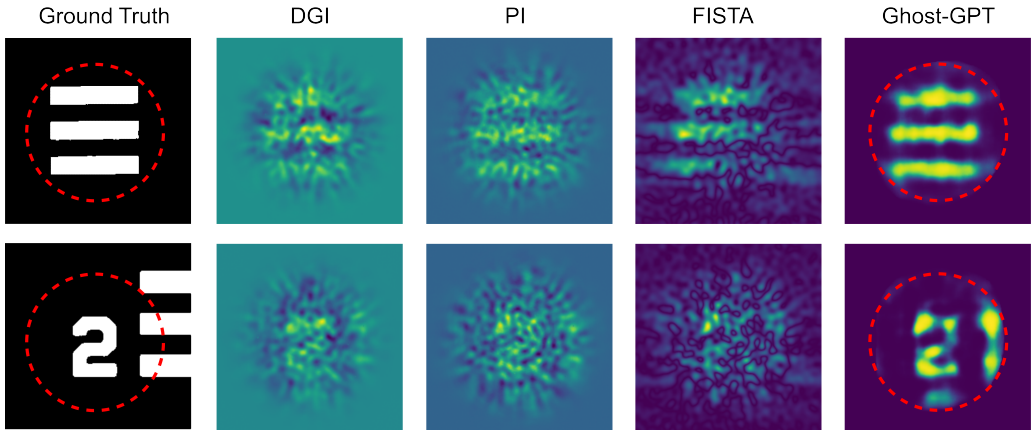

Figure 3: Reconstruction using experimental bucket measurements compared with classical algorithms and Ghost-GPT. (Top row): Striped lines correspond to group 0, element 4 (1.41 line pairs per millimeter). (Bottom Row): The number 2 corresponds to group 0, element 2 (1.12 lp/mm).

Table 1: MSE/SSIM for Experimental Targets with Classical Algorithms and Ghost-GPT

Experimental Target	DGI	PI	FISTA	Ghost-GPT (Ours)
Stripes	0.231/0.030	0.140/0.028	0.072/0.057	**0.045/0.604**
Number 2	0.204/0.036	0.138/0.028	0.084/0.064	**0.058/0.658**

To demonstrate the speed of our optical setup and reconstruction model, we experimentally demonstrated the ghost imaging of a dynamic scene. As shown in Figure 4, our system achieves video-rate ghost imaging at 60 FPS with high image fidelity for each frame.

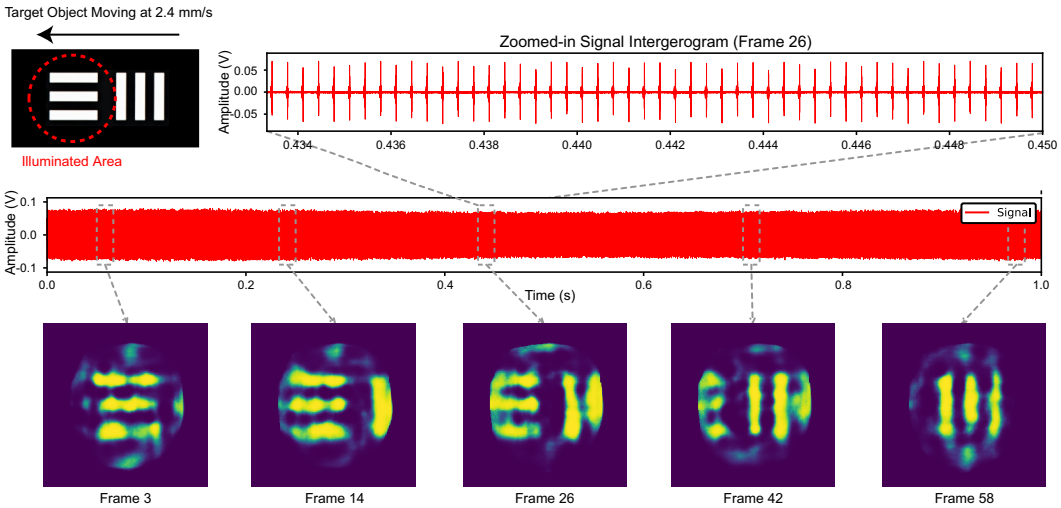

Figure 4: Video-rate image reconstruction of a moving target. A hyperspectral speckle pattern is projected onto a USAF1951 resolution target (Group 0, Element 4; 1.41 lp/mm), which is mounted on a motorized stage and translated horizontally at a speed of 2.4 mm/s.

References

[1] M. J. Padgett and R. W. Boyd, "An introduction to ghost imaging: quantum and classical," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 375, no. 2099, p. 20160233, 2017.

[2] D. Li, D. Yang, S. Sun, Y.-G. Li, L. Jiang, H.-Z. Lin, and W.-T. Liu, "Enhancing robustness of ghost imaging against environment noise via cross-correlation in time domain," *Optics Express*, vol. 29, no. 20, pp. 31068–31077, 2021.

[3] X. Liu, J. Shi, L. Sun, Y. Li, J. Fan, and G. Zeng, "Photon-limited single-pixel imaging," *Optics express*, vol. 28, no. 6, pp. 8132–8144, 2020.

[4] X. Liu, J. Shi, X. Wu, and G. Zeng, "Fast first-photon ghost imaging," *Scientific reports*, vol. 8, no. 1, p. 5012, 2018.

[5] F. Li, M. Zhao, Z. Tian, F. Willomitzer, and O. Cossairt, "Compressive ghost imaging through scattering media with deep learning," *Optics Express*, vol. 28, no. 12, pp. 17395–17408, 2020.

[6] L.-X. Lin, J. Cao, D. Zhou, H. Cui, and Q. Hao, "Ghost imaging through scattering medium by utilizing scattered light," *Optics Express*, vol. 30, no. 7, pp. 11243–11253, 2022.

[7] M. Don, "An introduction to computational ghost imaging with example code," *CCDC Army Research Laboratory: Aberdeen Proving Ground, MD, USA*, 2019.

[8] V. Kilic, T. D. Tran, and M. A. Foster, "Compressed sensing in photonics: tutorial," *Journal of the Optical Society of America B*, vol. 40, no. 1, pp. 28–52, 2022.

[9] I. Coddington, N. Newbury, and W. Swann, "Dual-comb spectroscopy," *Optica*, vol. 3, no. 4, pp. 414–426, 2016.

[10] C. Bao, M.-G. Suh, and K. Vahala, "Microresonator soliton dual-comb imaging," *Optica*, vol. 6, no. 9, pp. 1110–1116, 2019.

[11] E. Vicentini, Z. Wang, K. Van Gasse, T. W. Hänsch, and N. Picqué, "Dual-comb hyperspectral digital holography," *Nature Photonics*, vol. 15, no. 12, pp. 890–894, 2021.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[13] S. Jamil, M. Jalil Piran, and O.-J. Kwon, "A comprehensive survey of transformers for computer vision," *Drones*, vol. 7, no. 5, p. 287, 2023.

[14] S. Kashyap, S. Singh, and D. V. Singh, "Speech-to-speech translation using transformer neural network," in *International conference on soft computing for problem-solving*, pp. 813–826, Springer, 2023.

[15] F. Sufi, "Generative pre-trained transformer (gpt) in research: A systematic review on data augmentation," *Information*, vol. 15, no. 2, p. 99, 2024.

[16] B. Ghojogh and A. Ghodsi, "Attention Mechanism, Transformers, BERT, and GPT: Tutorial and Survey." working paper or preprint, Dec. 2020.

[17] P. Zhang, X. Dai, J. Yang, B. Xiao, L. Yuan, L. Zhang, and J. Gao, "Multi-scale vision longformer: A new vision transformer for high-resolution image encoding," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2998–3008, 2021.

[18] J. Maurício, I. Domingues, and J. Bernardino, "Comparing vision transformers and convolutional neural networks for image classification: A literature review," *Applied Sciences*, vol. 13, no. 9, p. 5521, 2023.