Centimeter-Wave Free-Space Neural Time-of-Flight Imaging – Supplemental Material

SEUNG-HWAN BAEK*, Princeton University
NOAH WALSH*, Princeton University
ILYA CHUGUNOV, Princeton University
ZHENG SHI, Princeton University
FELIX HEIDE, Princeton University

ACM Reference Format:
Seung-Hwan Baek, Noah Walsh, Ilya Chugunov, Zheng Shi, and Felix Heide. 2021. Centimeter-Wave Free-Space Neural Time-of-Flight Imaging – Supplemental Material . *ACM Trans. Graph.* 39, 4, Article 1 (July 2021), 16 pages. https://doi.org/http://dx.doi.org/10.1145/8888888.7777777

In this supplemental document, we provide additional details on the proposed method and further validation. Specifically, we describe

- Experimental Prototype Details.
- Electro-Optical Modulators.
- RF Driver and Demodulation Electronics.
- Detection.
- Detailed Derivation of GHz Modulation.
- Heterodyne ToF Mode.
- Network Architecture Details.
- Additional Experimental Results.
- Additional Synthetic Validation.

1 EXPERIMENTAL PROTOTYPE DETAILS

This section provides additional details on our acquisition and validation setups which we use to characterize the optical GHz modulation and demodulation.

1.1 Validation Setup

In order to validate the proposed all-optical GHz correlation ToF imaging, we build a prototype as shown in Fig. 1. One major difference of this validation setup as compared to our acquisition setup is that we do not emit the modulated light into the scene. Instead, we forward amplitude-modulated light directly to a photodiode and measure its magnitude. This allows us to test the GHz intensity modulation of each EOM module in isolation.

Validating GHz Amplitude Modulation. Our laser is a Quantum Gem 532 continuous wave TEM_{00} single mode at 532nm with a bandwidth of 30GHz. The laser light beam is coupled into an OZ Optics single mode high power optical fiber, which removes the higher order modal light and produces a uniform Gaussian beam at the output of the fiber, reducing the laser output intensity to $20-30\%$. The light passes through a convex lens, a HWP, and a concave lens, resulting in a narrow beam diameter at a specific linear polarization angle. Depending on the HWP angle, light splits into two separate paths to different GHz-modulation units. See Fig. 6. For the first unit, shown in the bottom right diagram of Fig. 1(a), we have a HWP that vertically polarizes incident light. Thus light passes through a polarizing beamsplitter without loss of intensity and its polarization state is modulated by a pair of a HWP and a QWP. This combined polarization and EOM modulation produces the GHz frequency modulated illumination as described in the main paper. The light from this GHz-modulation unit is then directed to a photodiode. This same principle is applied to the other GHz-modulation unit and we adjust the angle of the first HWP in order to validate the modulation quality of both EOMs.

Assessment of Modulation Envelopes. We measured the modulated amplitude of light by adjusting the HWP angle in front of the EOM. Fig. 2 shows these amplitude measurements for three different HWP angles of 11.25, 22.5, and 0 degrees. The placements of the HWP at 22.5/0 degrees cause frequency doubling near the minimum/maximum transmission, while the single frequency operation is obtained for 11.25 degrees. Reading the data directly from the photodiodes with a high-frequency oscilloscope, we confirm that the operation of our GHZ-modulation units aligns with the theoretical model.

2 ELECTRO-OPTICAL MODULATOR

GHz modulation on the optical carrier is produced by resonant electro-optic amplitude modulators (EOMs). Specifically, an EOM shifts the polarization of the light along its optical axis by an amount dependent on the applied voltage V. It achieves this by changing the phase between the perpendicular polarization components of light, represented by the Jones matrix

$$B(V) = \begin{bmatrix} e^{-i\Gamma(V)/2} & 0 \\ 0 & e^{i\Gamma(V)/2} \end{bmatrix}, \tag{1}$$

*Authors contributed equally to this work.

Authors' addresses: Seung-Hwan Baek, Princeton University; Noah Walsh, Princeton University; Ilya Chugunov, Princeton University; Zheng Shi, Princeton University; Felix Heide, Princeton University.

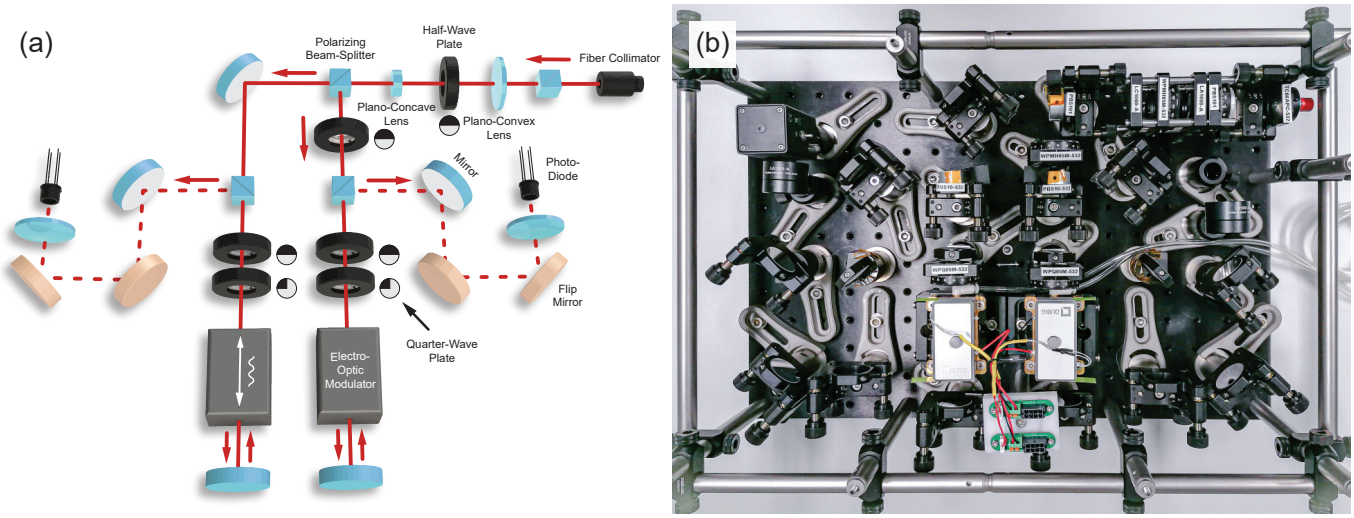

Fig. 1. (a) Illustrated schematic of our GHz-frequency validation setup. (b) Photo of assembled validation prototype, used to gather the GHz signal traces. This contains the EOMs that modulate the optical signal and the paths to the detectors.

where $\Gamma(V)$ is the voltage-dependent net birefringence of the EOM, taking the form

$$\Gamma(V) = \frac{2\pi}{\lambda}\Delta n^3 r V = \pi V / V_\pi. \tag{2}$$

Here Δn is the birefringence, the difference between the ordinary and extaordinary refractive indices of the uniaxial crystal, and r is the element corresponding to index $[3, 3]$ of the electro-optic tensor for the Lithium Niobate crystal, and V_π is the half wave voltage. Note that Eq.8. in the main paper is the simplified version of Eq. (1) & (2), assuming $V_\pi = \pi$.

2.1 Resonance Modulator

A resonant standing wave electro optic modulator is comprised of a housing containing an electro optic crystalline material. This housing forms a resonant cavity that produces standing waves of a modulating frequency. A resonant circuit is used to achieve true impedance matching and reduce the required drive voltage. The most basic type involves the LC tank, wherein a low-loss inductor and modulator crystal are used to form a series resonant circuit that acts as a small resistor while at resonance, whose value depends on the losses from the inductor. The 50 Ohm driving impedance is matched to the resistance with a transformer, and this impedance matching to the source along with low loss components results in a voltage across the capacitor ten times greater than the input voltage. This is a result of the energy storing properties of the resonant circuit, which allows further reduction of half-wave voltages as compared to a broadband modulator.

The resonator is power-limited, and can achieve only a narrow modulation band. *This means we can only sample our frequency 7.15 GHz with a small bandwidth, around 20 MHz. This is opposed to EOMs in the MHz and below frequency range, which have a much wider bandwidth, such as a MHz modulator that can go from 1-100 MHz.*

RF properties	Emission	Detection	Unit
Resonance frequency (f_{set})	7150	7150	MHz
Bandwidth (δ v)	17.6	20.4	MHz

Table 1. Resonant electro-optic amplitude modulator RF specifications for the emission (second column) and the detection (third column).

2.2 Amplitude Modulation with EOM

We modulate the amplitude of light with a EOM and polarizing optics shown in Fig. 6. The input light is elliptically polarized by a HWP and QWP mounted at 11.25 and 45 degrees, respectively. It is then reflected by a mirror which is half a wavelength away from the EOM. This light is modulated very strongly at the RF driving frequency, and, returning from the modulator, passes through the QWP and HWP at the same angles to become vertically polarized. The full derivation, with Jones Matrices, is shown in the main paper. This light is now amplitude modulated at the RF frequency. The special cases of frequency doubling occur when the HWP is shifted to 0 and 22.5 degrees.

2.3 EOM Performance

The custom EOMs we are using in our work are resonant, meaning they contain an LC tank resonator circuit that limits the frequency range to a very narrow band. Measured RF properties and optical properties of the two EOMs are provided in Tab. 1 and 2. Here, a resonant frequency is one which, due to its proximity to a natural frequency of the system, leads to a increase in signal amplitude.

2.4 Temperature Control

Temperature control is a critical component for the EOM modulators. We use chillers (Solid State Cooling Systems) to cool the modulators down to operate at maximum efficiency at about 27-29 degrees Celsius.

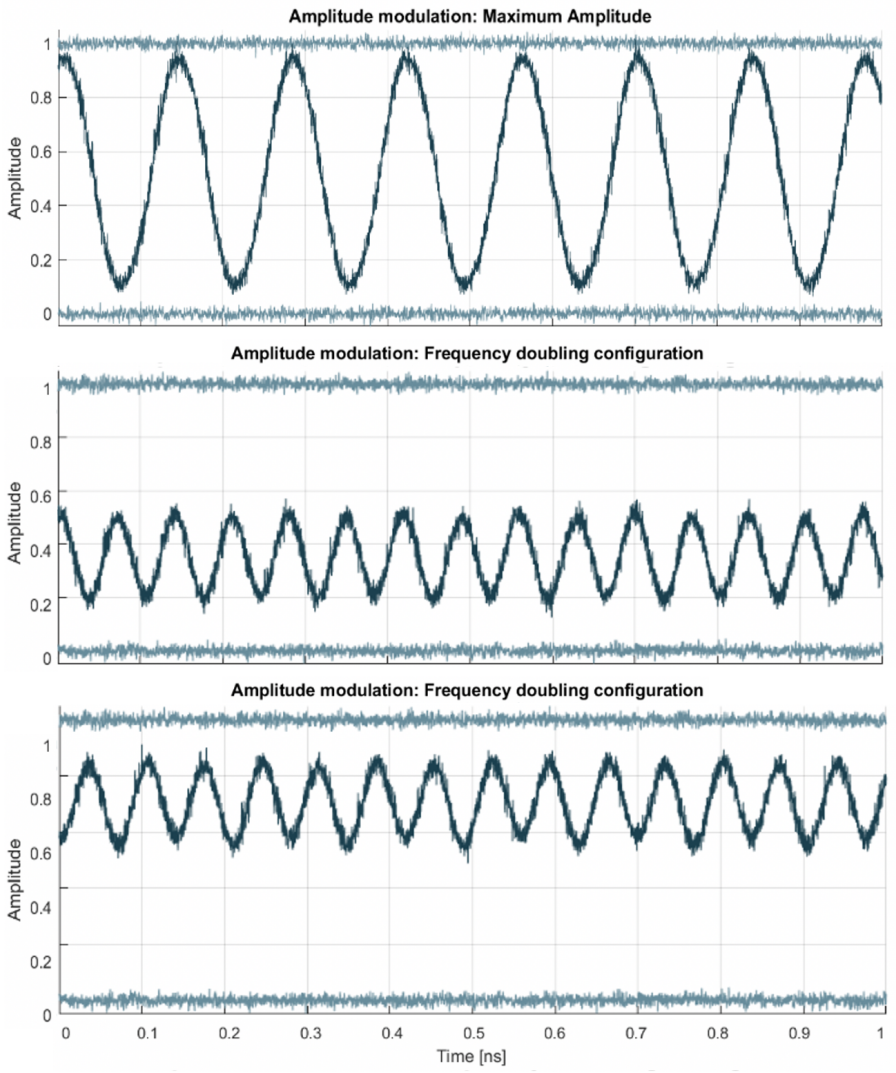

Fig. 2. Frequency traces. The trace at the top is modulation of the fundamental frequency. The middle trace is frequency doubling modulation near minimum transmission. The bottom trace is frequency doubling modulation near maximum transmission. Notice that for both frequency doubled signals, their amplitude is half that of the fundamental frequency.

Optical properties	Value	Unit
Aperture	Φ 1.8	mm
Wavefront distortion (@ 633 nm)	$\lambda/4$	nm
Bandwidth (δ v)	20.4	MHz

Table 2. Resonant electro-optic amplitude modulator EOM specifications. Note that both EOMs for the emission and the detection have identical values.

3 RF DRIVER AND DEMODULATION ELECTRONICS

We use two custom RF drivers with a high-frequency DDS to produce the RF signals driving the EOMs, as well as shifting the phase. This subsection describes operational and implementation details on this RF driver for the EOMs. We connect the EOM to the RF driver output via a low-attenuation SMA cable. We also use *frequency locking* to ensure that modulation frequencies of the RF driver and the EOM's resonance frequency are matched.

3.1 Reference Clock

We a function generator (Siglent SDG2042X) to produce an external 10 MHz square wave reference clock. This allows the two RF drivers for the emission and the detection EOMs to be accurately phase synced.

3.2 RF Demodulation

We implement the all-optical demodulation of return signals using a detection GHz-modulation unit. For fair comparison with

RF-based GHz ToF imaging, we also include the hardware necessary for demodulation electrically aftr photodetection. To this end, we implement an I/Q demodulator, which computes the I'-in phase component and Q'- quadrature component, with 90 degrees reference phase difference. By calculating the ratio of I and Q, we can determine $R = \sqrt{I^2 + Q^2}$, the amplitude of the signal, and $\theta = \arctan(Q/I)$ the phase shift of the return signals. This works via a lock-in amplifier (LIA) on the RF driver. Note that before conducting the I/Q

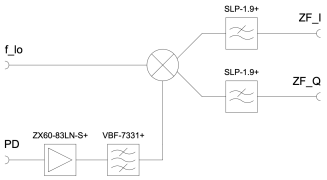

Fig. 3. RF Mixer diagram. Describes the I/Q demodulator circuit where the photodetected signal is amplified, filtered, then mixed with the local oscillator signal to produce the in-phase (I) and quadrature (Q) components.

demodulation, we filter out the signals from the photodetector with a low noise amplifier (Mini-Circuits ZX60-83LN-S+) and a coaxial bandpass filter (Mini-Circuits VBF-7331+) to filter the signal. See Fig. 3. This is then mixed with the local oscillator signal f_{lo} to produce two frequencies, the sum of the two frequencies ($f_{lo} + f_{PD}$) and the difference of them ($f_{lo} + f_{PD}$). This sum is then low-pass filtered, in this case with a 10KHz low-pass and 10ms RC filter. The signal is split along two paths, one with a phase change of 90 degrees, and the other in phase. These signals pass through the low pass filters SLP-1.9+ (Mini-Circuits), with bandwidth DC to 1.9 MHz. Then it goes out to ZF_I and ZF_Q, then to the detection electronics.

4 DETECTION

After the signal is emitted to the scene and returns, we concern ourselves with the task of detection. In general, the returned beam has a phase shift due to the time delay τ, and a new amplitude and bias. To recover these parameters, we first collect as much of the returned light as we can through focusing and collimating optics. We then pass this collimated light through a second EOM on the detector side, that acts as a reference signal demodulator, correlating it with the signal beam to produce a homodyne DC signal. The resultant demodulated signals are focused onto a photodiode to be converted to a photocurrent. Our APD effectively picks up a continuous DC signal and we further filter it out with a BNC 10kHz lowpass filter, and a resistor capacitor (RC) low-pass integrator circuit. Lastly, it is passed to an analog-digital-converter (LabJack T7) to sample the signal at up to 24 ksamples per second. We then computationally recover phase, amplitude and bias. Details of each process is described as follows.

4.1 Photodiodes

Avalanche Photodiode for our AMCW ToF. We use an APD (Thorlabs APD440A) that has significantly higher gain than a typical PIN diode to measure a few μW returned from the scene demodulated

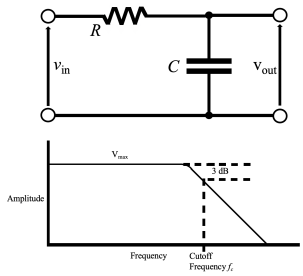

Fig. 4. RF Integrator and Low Pass filter, essential for removing higher frequency noise from the homodyne signals to produce a smooth DC value. Circuit diagram consists of a resistor and capacitor (top). Spectral plot showing how higher frequencies are removed above the cutoff frequency (bottom).

by the second EOM. The APD reads out the continuous DC signal after optical correlation.

GHz Photodetector for RF-based ToF. For comparison with post-photoconversion methods, we use a high-speed 12 GHz photodetector (EOT GaAs PIN Detector ET-4000) with modulation after amplification and sensing. Our GHz photodetector comes with internal bias supply consisting of lithium cells. It has an active area diameter of 60 μm, and an acceptance angle of 15deg requiring careful optical alignment.

Integrating Sphere for Laser Calibration. An additional photodetector is an integrating sphere that is placed at the second beamsplitter after the emitting path, before passing through the third beamsplitter and into the scene. See Fig. 4 in the main paper. The sphere measures power fluctuations from our laser, which is used to normalize all the returned signals incident on the previous three detectors. It is used for both our all-optical amplitude modulated-demodulated method and the analog RF I/Q demodulator method, and the interference measurements. The output from the sphere is sent into a digital power meter (Thorlabs PM100D), which is mapped to a voltage value between -0.3 and 2 V, sampled at the same rate in sync with the returned signal detection measurements. The placement of the sphere after the modulation of the input is acceptable as the laser intensity bias is significantly greater than the applied modulation and the modulation is fast enough that it averages out in the sphere's integration. Then the sphere only measures the intensity fluctuation from the laser input.

4.2 RF Integrator Circuit

The RF Integrator circuit consists of a resistor and capacitor in a low-pass filter configuration. For a resistor R, capacitor C, the cutoff frequency f_0 at which signals are attenuated by 3 dB is

$$f_0 = \frac{1}{2\pi RC}, \tag{3}$$

The values of resistance and capacitance form the RC time constant which determine the rise time of the circuit, or the amount of time it takes the capacitor plates to charge.

A simple passive Low Pass Filter or LPF, can be made by connecting together in series a single resistor with a single capacitor

as shown in the top of Fig. 4. In this type of filter arrangement the input signal is applied to the series combination, both the resistor and capacitor together, but the output signal is taken across the capacitor only.

This delay, if long compared to the signal being input will act as an integrator and low pass filter. So for integration, we only taking into accurate high frequencies $\omega >> \frac{1}{RC}$ so that the capacitor does not have time to charge and discharge completely, blocking higher frequencies past a cutoff as shown in the bottom of Fig. 4. Then its voltage is small enough that the input voltage is equal to the voltage across the resistor:

$$v_{out} = \frac{1}{RC} v_{in} dt. \tag{4}$$

We used $C = 50$ nF and $R = 10$ kOhms, such that the output is into the DAQ (see following subsection) is significantly smoothed.

4.3 Data Acquisition

For data acquisition, we use a multifunction DAQ with ethernet and USB (LabJack T7) for high speed flexible analog to digital conversion of the data. With 16-bit to 24-bit analog inputs, it allows for high speed data acquisition to a capture PC. It has 14 analog inputs built in, with 16-bit high-speed ADC (up to 100k samples/s), and contains single-ended Inputs, 14, or differential Inputs (7) and analog input ranges: ±10V, ±1V, ±0.1V and ±0.01V. All analog input features are software programmable by configuring the analog input registers, and high speed sampling configurable by using Stream Mode, and we use the 54kS/s mode.

4.4 Galvo Control

We use a scanning Galvanometer Mirror Positioning System (Thorlabs GVS012 Galvo), which is designed for laser beam steering applications with a beam diameter of <10 mm. It has a dual-axis galvo motor and mirror assembly, associated driver cards, and driver card heatsinks. Our controller consists of a microcontroller (Arduino) connected to 2 DACs to produce analog output for moving the galvos.

4.5 Photon Efficiency

To maximize imagine and ranging quality, a sensor module needs to be able to covert as many photons from the imaged scene that hit its surface to photocurrent as efficiently and as fast as possible. Observing our setup, we have 30 mW of light being produced by the laser, which is coupled into a high power single mode optical fiber. Achieving the best coupling possible, along with loses in the fiber, there is an output efficency of 30%, 10 mW which is set as input to our optical prototype. We then have losses in the optical system due to splitting of light paths at polarizing and non-polarizing beamsplitters, as well as losses in transmission through interreflections. Then the output to the scene for this input is measured as approximately 3 mW, with a loss of one half due to each beamsplitter plus other losses. Then to determine the maximum energy converted by the detector, we multiply the input to the scene by exposure time on the detector. We can assume the maximum return power is $P_m = 3$ mW for a scene with a specular mirror at normal incidence to the beam. For our analog-digital-converter (ADC), it samples at a rate of 54000 samples/s, with a resolution of 14 bits. Then for 50 samples for 16 phases, we integrate over the exposure time of $t_e = 13$ ms. Then the maximum energy E_m incident on the sensor is

$$E_m = P_m \cdot t_e = 39\ \mu\text{J}.$$

The maximum number of photons is determined by dividing this maximum energy by the energy of a single photon of wavelength $\lambda = 532$ nm light, which is $E_p = \frac{hc}{\lambda}$, where the Planck constant $h = 6.626 \cdot 10^{-34}$ J/s is a quantum of electromagnetic energy, and the speed of light $c = 2.998 \cdot 10^8$ m/s. Then the energy of a 532 nm photon is

$$E_p = \frac{hc}{\lambda} = \frac{6.626 \cdot 10^{-34} \cdot 2.998 \cdot 10^8}{532 \cdot 10^{-9}} = 3.734 \cdot 10^{-19}\ \text{J}.$$

Then the total number of emitted photons N_p that could be incident on the detector is

$$N_p = \frac{E_m}{E_p} = 1.0445 \cdot 10^{14}\ \text{photons}.$$

We calculate the number of photons returned from a diffuse scene set a distance $R = 1$ m away from the detector. We assume a uniform hemisphere of photons for a completely diffuse reflection. The second modulator has a $r = 1$ mm radius aperture, so the number of photons N_d that fall into this 1 mm radius of the hemisphere on the modulator are the actual number of detected photons. Then the total curved surface area CSA of the hemisphere draws out $CSA = 2\pi R^2 = 2\pi\ \text{m}^2$, and then the capped surface area SA is equal to $SA = 2\pi Rh$ where the height is $h = R - \sqrt{R^2 - r^2}$, so we calculate

$$SA = 2\pi R(R - \sqrt{R^2 - r^2}) = 3.14 \cdot 10^{-6}\ \text{m}^2.$$

Then the ratio between the SA and CSA is

$$\frac{SA}{CSA} = \frac{3.14 \cdot 10^{-6}}{2\pi} = 0.5 \cdot 10^{-6}.$$

This is also the ratio of detected photons from the scene to emitted photons to the scene. so then we multiply this ratio by the total number of reflected photons N_p to get the true number of photons incident on the detector N_d from a diffuse reflection, which comes out to

$$N_d = N_p \cdot \frac{SA}{CSA} = 5.222 \cdot 10^7\ \text{photons}.$$

This is the number of photons we expect to see on our detector for a diffuse reflection. As such, an upper bound of the detection photon efficiency is 8%. We experimentally measure 5% efficiency in the proposed free space setup which fares close to the upper bound. Please note that we have not added additional collection optics to the proposed coaxial setup.

We note that our system has many optical elements, resulting in energy loss of beam due to reflections and refractions between the elements. Miniaturization of the benchtop optical setup and the optical path in a mobile-phone form factor could mitigate this problem via optical engineering.

4.6 Interferometric Capture

We also compare the proposed method to interferometric depth acquisition. Interference techniques can be used for studying light transport [Gkioulekas et al. 2015], for example computational imaging systems such as optical coherence tomography (OCT) use interferometry to produce decompositions of light transport in small

scene volumes of a few centimeters. To compare against the proposed system at a longer standoff distance, we use conventional Michelson interferometer.Here, an incident light beam is split into two identical beams by a beamsplitter, and each new beam travels down a different path, and reflects off a mirror at an independent distance, shown in Fig. 5. This signals travel back down their respective paths and recombine at the beamsplitter, and interfere with each other as they travel down toward the detector, forming a circular interference pattern.

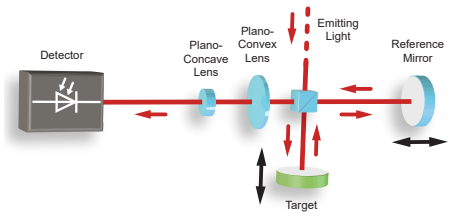

Fig. 5. Michelson interferometry incorporated into our setup. The reference mirror is adjusted for each point in the scene to reconstruct submicron details.

We incorporate interferometry in our setup shown in Fig. 5. We have a beamsplitter placed between the scene and the emission path, which allows for the light to be split along the left direction. A mirror on a piezoelectric stage is placed behind this side of the BS, and now that serves as the reference arm. The stage can travel a total of 21μm, and has a theoretical resolution of 76 nm, small enough to see interference in the 532 nm light. The object of interest in the scene, such as a mirror for specular reflection, is placed on a motion stage at 60 cm distance. This motion stage has a theoretical resolution of 50 nm, allowing for submicron depth information to be measured in the interference pattern, which is later recovered. This interference plot can be used to study properties of the scene when converted to amplitude and phase plots.

5 DETAILED DERIVATION OF GHZ MODULATION

This section provides detailed derivations of our two-pass GHz modulations extended from the main paper. For completeness, we describe the entire derivation. See Fig. 6 in the main paper for corresponding system schematics.

Light enters a polarizing beam splitter turning light into vertical linear polarization as

$$E_0 = A\begin{bmatrix}0\\1\end{bmatrix}, \tag{5}$$

where A is the amplitude of the incident light. The polarization state of the light is then modulated by a HWP and a QWP followed by a EOM at a given voltage V as

$$E_1 = B(V)Q(\theta_q)H(\theta_h)E_0, \tag{6}$$

where the Jones matrices of the HWP and the QWP oriented at angle θ_h and θ_q are defined as

$$H(\theta_h) = e^{-i\pi/2}\begin{bmatrix}\cos^2\theta_h - \sin^2\theta_h & 2\cos\theta_h\sin\theta_h\\ 2\cos\theta_h\sin\theta_h & \sin^2\theta_h - \cos^2\theta_h\end{bmatrix},$$
$$Q(\theta_q) = e^{-i\pi/4}\begin{bmatrix}\cos^2\theta_q + i\sin^2\theta_q & (1-i)\cos\theta_q\sin\theta_q\\ (1-i)\cos\theta_q\sin\theta_q & \sin^2\theta_q + i\cos^2\theta_q\end{bmatrix}.$$

The exitant light from the EOM then propagates in free-space by the distance corresponding to the half of the modulation wavelength c/ω where a mirror is placed, resulting in the change of Jones vector as

$$E_2 = ME_1, \tag{7}$$

where M is the Jones matrix of a mirror

$$M = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}. \tag{8}$$

Light travels again back to the EOM, the QWP, and the HWP and the PBS picks up the vertical linear polarization component of the light. Setting the HWP and the QWP angles as $\theta_q = 11.25°$ $\theta_q = 45°$, we obtain the output light

$$\begin{aligned}E_3 &= L_hH(-\theta_h)Q(-\theta_q)E_2\\ &= L_hH(-\theta_h)Q(-\theta_q)B(V)MB(V)Q(\theta_q)H(\theta_h)E_0\\ &= \begin{bmatrix}\frac{(1+i)e^{-iV}(i+e^{2iV})}{2\sqrt{2}} & \frac{e^{-iV}((1+i)-(1-i)e^{2iV})}{2\sqrt{2}}\\ 0 & 0\end{bmatrix}E_0\\ &= \begin{bmatrix}\frac{i(\cos V+\sin V)}{\sqrt{2}} & \frac{i(\cos V-\sin V)}{\sqrt{2}}\\ 0 & 0\end{bmatrix}E_0\\ &= A\begin{bmatrix}\frac{i(\cos V-\sin V)}{\sqrt{2}}\\ 0\end{bmatrix},\end{aligned} \tag{9}$$

where L_h is the Jones matrix of the horizontal linear polarizer

$$L_h = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}. \tag{10}$$

We apply the squared magnitude to E_3, resulting in the modulated intensity $I(V)$ as

$$I(V) = |E_3|^2 = \frac{A^2}{2}(1-\sin 2V). \tag{11}$$

Eq. (11) indicates that the output intensity of light is a function of the voltage V applied to the EOM. As we supply a time-varying sinusoidal voltage to the EOM, we arrive at the time-varying intensity-modulated light as

$$\begin{aligned}I(t) &= \frac{A^2}{2}(1-\sin(2\eta\sin(\omega t-\phi)))\\ &\approx \frac{A^2}{2}(1-2\eta\sin(\omega t-\phi)).\end{aligned} \tag{12}$$

The last approximation is based on the Taylor expansion assuming the small modulation power η. The applied voltage to the EOM has GHz modulation frequency ω, enabling effective all-optical GHz modulation of light intensity. We refer to the Supplemental Material for additional detail.

Eq. (12) describes the high-frequency intensity modulation realized by our free-space optical setup shown in Fig. 6. This optical

configuration serves as a *building block for both illumination and detection modules* in our imaging system. In the illumination module, we input continuous laser light into the EOM, resulting in sinusoidally intensity-modulated light emitted into the scene as p. For the detection module, the returned amplitude-modulated light $\tilde{p}$ from the scene is demodulated by an additional intensity modulation with the reference signal r and we *optically* multiply r and $\tilde{p}$ before integration on the detector.

Double-Frequency Modulation. Even though the voltage modulation frequency ω is limited to a narrow modulation band in our resonant EOM, we can modulate at the double frequency of 2ω by adjusting the angle of the HWP, θ_h, in front of the EOM. While doubling the frequency of the optical carrier is well known in optics, we note that the proposed frequency doubling of the RF intensity modulation is novel. In the original operating mode, we set θ_h as $11.25°$ resulting in the intensity modulation at ω. For frequency doubling, we rotate the HWP to $\theta_h = 22.5°$. To derive the modulation behavior, we rely on the same Jones calculus from above. Specifically, changing θ_h in the polarization in the system of Jones matrices results in the output light E_3 as

$$\begin{aligned} E_3 &= L_h H(-\theta_h) Q(-\theta_q) E_2 \\ &= L_h H(-\theta_h) Q(-\theta_q) B(V) M B(V) Q(\theta_q) H(\theta_h) E_0 \\ &= \begin{bmatrix} \frac{ie^{-iV}(i+e^{2iV})}{2} & -\frac{e^{-iV}(-1+e^{2iV})}{2} \\ 0 & 0 \end{bmatrix} E_0 \\ &= \begin{bmatrix} i\cos V & -i\sin V \\ 0 & 0 \end{bmatrix} E_0 \\ &= A \begin{bmatrix} -i\sin V \\ 0 \end{bmatrix}. \end{aligned} \tag{13}$$

The intensity $I(V)$ is the magnitude square of E_3 as

$$I(t) = |E_3|^2 = \frac{A^2}{2}(1 - \cos(2V)). \tag{14}$$

Note that the difference of Eq. (14) with Eq. (11) is that we have cos() instead of sin(). This single difference enables us to arrive at the intensity modulation at double frequency. After applying the time-varying voltage modulation, the time-varying intensity of the output light is

$$\begin{aligned} I(t) &= \frac{A^2}{2}(1 - \cos(2\eta \sin(\omega t - \phi))) \\ &\approx \frac{A^2}{2}\eta^2 \sin^2(\omega t - \phi) \\ &= \frac{A^2}{4}\eta^2(1 - \cos(2\omega t - 2\phi)) \end{aligned} \tag{15}$$

Note that we use the same Taylor expansion with small modulation power η in the second approximation. Eq. (15) shows that we can obtain doubled frequency modulation of 2ω with reduced amplitude by the factor of four compared to the single-frequency mode at ω – only by changing the polarization optics instead of the electro-optical modulation itself.

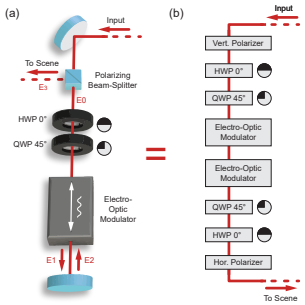

Fig. 6. Illustration of our design and equivalent schematic. The left is our EOM setup that modulates and demodulates the optical carrier with amplitude modulation. The right is a schematic of the unwrapped setup with a straight path and twice as many components.

6 HETERODYNE TOF MODE

In our work, we use the same frequency for both emission and detection paths in a *homodyne* configuration. The case where we have differing modulation and demodulation frequencies, that is $\omega_p \neq \omega_r$, is known as a *heterodyne* setup. If we assume that there is a relatively small frequency shift between the two, i.e. $\omega_r = \omega_p + \omega_\delta$, integrating the correlated equation from the main paper we find that the correlation can be approximated as

$$\begin{aligned} C_\psi(\rho) &= \int_\rho^{\rho+T} \tilde{p}(t-\tau) r(t)\, \mathrm{d}t \\ &= \frac{\tilde{\alpha}}{2}\int_\rho^{\rho+T} \cos\left((\omega_p - \omega_r)t - \phi + \psi\right)\, \mathrm{d}t + \\ &\quad \underbrace{\frac{\tilde{\alpha}}{2}\int_\rho^{\rho+T} \cos\left((\omega_p + \omega_r)t - \phi + \psi\right)\, \mathrm{d}t}_{\approx 0} + \\ &\quad \underbrace{\tilde{\beta}\int_\rho^{\rho+T} \cos(\omega_r t + \psi)\, \mathrm{d}t}_{\approx 0} + TK \\ C_\psi(\rho) &\approx \frac{\tilde{a}}{2}\cos\left(\omega_\delta - \phi + \psi\right) + TK. \end{aligned} \tag{16}$$

where once again we use the property that for $T \gg \omega_r$ and $\omega_r \gg \omega_\delta$, this integration acts as a low-pass filter, eliminating the high frequency terms. Unlike the homodyne case, however, the resultant correlation function $C_\psi(\rho)$ now depends on ρ, varying sinusoidally with frequency ω_δ. We can thus estimate the value of the phase shift ϕ and calculate depth in a similar fashion to the homodyne setup, via Fourier analysis of $C_\psi(\rho)$ measurements. However, rather than acquiring values at different offsets ψ, we achieve this by sampling values at varying ρ, as the function's phase naturally shifts over time.

7 FINE-TUNING DETAILS

We fine-tune our neural unwrapping method using a training dataset of real-world scenes. To acquire pseudo ground truth wrapping

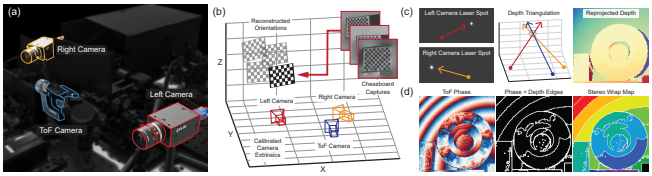

Fig. 7. (a) We augment our ToF imager with a pair of stereo cameras to acquire fine-tuning data. (b) Checkerboard-based geometric calibration provides all the geometric transformations between the three cameras. (c) For the training-set scenes, we detect laser spots in left and right camera views from which depth is computed via triangulation with respect to the ToF imager. (d) The reconstructed stereo depth combined with the measured ToF phases enable us to prepare dataset for fine-tuning.

counts, we mount two color cameras (FLIR Grasshopper3 GS3-U3-32S4C) with lenses (Tamron 8mm 1/1.8inch) to the GHz ToF rig. See Fig. 7. Geometric calibration is conducted using the checkerboard-based method [Zhang 2000]. We use these camera views to triangulate the position of the ToF laser spot in 3D space as we scan the scene. We then re-project the left-right depth, which has the largest baseline, to the ToF camera coordinate system. Combining the depth edges from stereo with discontinuities in the ToF phase measurements we can form watershed segments of contiguoes wrap zones, see Fig. 7 (d). The median stereo depth of each of these zones gives a robust estimate of its wrap count, which allows us to generate a ground truth wrap map. We use this stereo-assisted ground truth to help bridge the domain gap between the simulated scene data and real captures. We perform network fine-tuning on diverse captured scenes with stereo measurements, which are withheld from the experimental result section. Our pre-trained network is supervised with L1 and Cross Entropy loss as used during training, as well as a gradient loss based on the raw phase measurement, where we mask depth and phase discontinuous regions to avoid the undue influence of their non-smooth phase gradients. This process helps us recover fine scene details that are otherwise not present in the simulated dataset scenes.

8 NETWORK ARCHITECTURE DETAILS

In this section, we provide additional details on the network architecture. Our phase unwrapping network uses a Fast SCNN [Poudel et al. 2019]-like architecture with 3 streams. We concatenate the Fourier encoded phase and phase edges as model input. The input is first downsampled to extract high resolution features. It is then feed into 3 bottleneck residual blocks for low resolution feature extraction. In parallel to a classifier, an auxiliary stream is applied in parallel to boost the system accuracy further. All feature fusions are done by addition for fast computation. The network architecture is described explicitly in Tab. **??**.

9 ADDITIONAL SYNTHETIC VALIDATION

9.1 Background on CRT Unwrapping

Many multi-frequency phase unwrapping methods for ToF fundamentally either weigh Euclidean division candidates [Bioucas-Dias et al. 2009; Droeschel et al. 2010; Lawin et al. 2016] or use CRT-like frequency-space lookup tables [Gupta et al. 2015] to estimate wrap counts. We thus follow with a description of CRT-based phase unwrapping as it gives insight into the logic behind, and failure cases of, such methods.

While there exist single-frequency phase unwrapping [Herráez et al. 2002] approaches, these suffer from problems of reference ambiguity; if there is no data with zero wraps in the measurement, where do you start unwrapping? If there are phase discontinuities, how many wraps exist between them? CRT-based unwrapping addresses this issue by asking for a minimum of two measurements for any point, at coprime modulation frequencies ω_1 and ω_2 (i.e $gcd(\omega_1, \omega_2) = 1$, where $gcd(\cdot)$ is the greatest common divisor). This means if our wrapped measurements for these frequencies are $\hat{\phi}_1$ and $\hat{\phi}_2$ we have a system of equations

$$\begin{aligned} \phi_1 &= \hat{\phi}_1 + 2\pi n_1, \quad n_1 \in \mathbb{N}, \\ \phi_1 &= k\phi_2 = k(\hat{\phi}_2 + 2\pi n_2, \quad n_2) \in \mathbb{N}, \\ \rightarrow (\hat{\phi}_1 - k\hat{\phi}_2) &- 2\pi(n_1 - kn_2) = 0. \end{aligned} \tag{17}$$

Where by the Chinese Remainder Theorem [Pei et al. 1996], for which the method is named, we have that this last equation admits a single set of solutions (n_1, n_2), which allow us to fully disambiguate ϕ_1. Here k is a constant factor which allows us to convert from the space of ω_2 to ω_1. In practice, given finite precision, (17) is seldom equal to exactly zero and so we instead seek to solve

$$\text{minimize}_{n_1,n_2} \left((\hat{\phi}_1 - k\hat{\phi}_2) - 2\pi(n_1 - kn_2) \right)^2, \tag{18}$$

for some set of candidates (n_1, n_2). Checking every integer combination proves computationally prohibitive, so we limit ourselves to candidates within some range [min_wrap, max_wrap] defined by the expected min and max depths in our scene. Even this leads to $O(n^2)$ combinations, so we can further reduce our set to frequency-feasible values. That is if $\omega_1 \approx 2\omega_2$ we test candidates $(n_1, floor(n_1/2) \pm 1)$, where we expect that $\hat{\phi}_2$ should have approximately half the wraps of $\hat{\phi}_1$; this greatly reduces the complexity to $O(n)$.

9.2 Noiseless Classical Results

To validate that our implementations of the classical baselines were in good faith, we tested their performance of noise-free simulated data, and display the qualitative results below in Fig. 8.

We observe that CRT [Xia and Wang 2007] achieves a perfect reconstruction of the original depth map used for simulation, as the 7.15GHz and 14.32GHz frequencies used for simulation are perfectly coprime in the simulated range of scene distances (0-2.5m). This means every pair of frequencies, unaltered by noise, is binned to the correct wrap count. The Phasor [Gupta et al. 2015] approach, with micro-shifted simulated frequencies at 7.15GHz and 7.16GHz, faithfully reconstructs the majority of the scene, but suffers from a periodic wrapping error. This is as the frequencies are not *optimal candidate frequencies* and thus leading to poor solutions when used in conjunction with the provided phase lookup table. Note that optimal frequencies would lie outside the 20MHz band of our modulators, which makes the two frequencies at the ends of the modulation band the local optima. KDE [Lawin et al. 2016] suffers minor scale issues and edge artifacts; likely due to the Gaussian

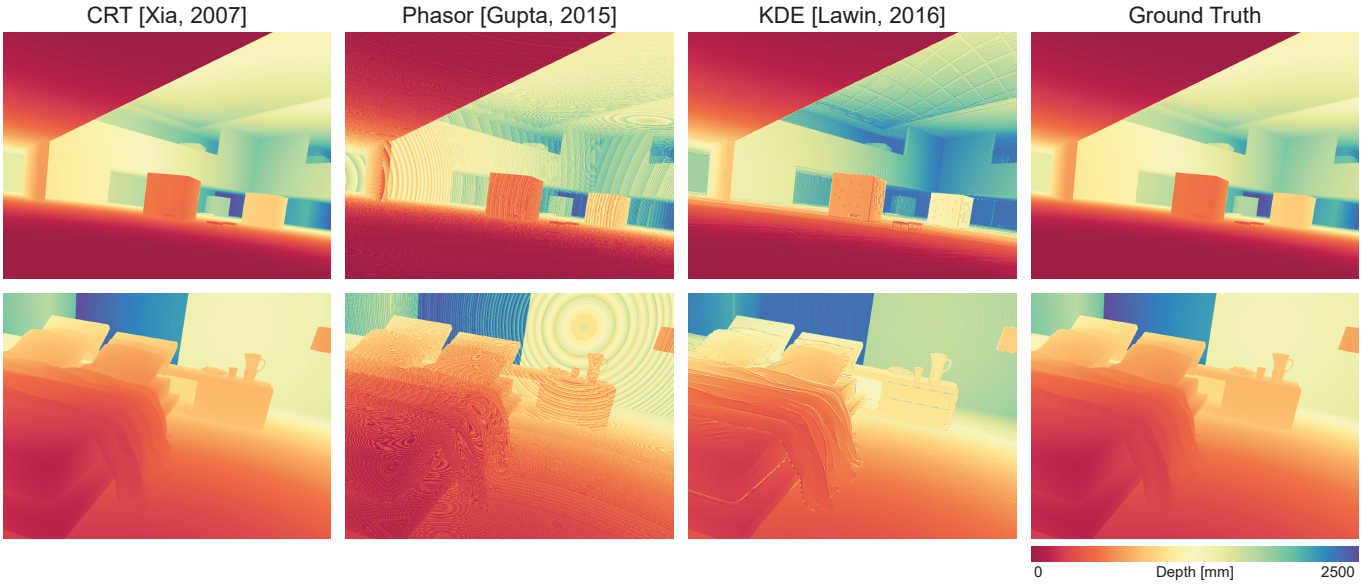

Fig. 8. Classical reconstruction results for synthetic test data simulated with no added Poisson-Gaussian noise.

kernel averaging over multiple closely-packed wraps, as the method was tuned for 16-120MHz frequencies. Nonetheless, without the perturbation of Poisson-Gaussian noise, it correctly reconstructs the majority of the scene.

9.3 Ablation results

In this section we provide network ablation results in Fig. 9 below. In the quantitative results we confirm our qualitative finds, as summarized in Tab. 4, seeing a drastic reduction in wrapping artifacts when we augment network input from raw phase to Fourier encoded measurements. Without the amplitude channel we see a visible degradation in object borders and spatial consistency, as seen on the edges of the blanket in the *bed* example. *Cross-Entropy Loss Only* and *Proposed* results look visually similar, as they are numerically so, yet we note a definite reduction in surface fluctuations when we add the combine ℓ_1 loss term.

9.4 Additional Results

In this section we showcase additional qualitative reconstruction results for all comparisons as listed in Tab. 5, as well as our proposed neural unwrapping method.

These further support claims on spatial consistency and smoothness as detailed in the main text, and showcase the successive improvements in performance between the baseline learned methods. Visually we see an increase in spatial diversity between *One-Step* [Wang et al. 2019] and *U-Net* [Ronneberger et al. 2015], with the latter predicting a wider range of classes. This could possibly stem from *One-Step* being primarily tailored towards low wrap count problems. While it produces visually consistent depth, the skip-connection rich network in *Deep-ToF* [Su et al. 2018] appears to have large global wrap errors, hallucinating measurements rather than unwrapping them. While the deepv3 network in the *Rapid.* [Zhang et al. 2019] approach is able to reconstruct small features, our proposed method achieves even finer reconstructed details with a significantly more compact computational footprint (≈8mil parameters vs 100mil+ parameters in the *Rapid.* architecture.).

10 ADDITIONAL EXPERIMENTAL VALIDATION

10.1 Comparison with the Unwrapping Methods

As outlined in the experiments shown in Tab. 5, we test our neural unwrapping method on real data and compare to the next-best learning method, *Rapid.* [Zhang et al. 2019] as well as analytic phasor unwrapping [Gupta et al. 2015]. Fig. 13 shows that our neural-unwrapping method recovers both scale and geometric details with clear object boundaries while *Rapid.* results in a significant loss of depth features and edge artifacts inside object boundaries. The lookup-based phasor method fails to handle the large number of wraps from the GHz modulation frequency.

10.2 Comparison with the MHz AMCW ToF Camera

Our method achieves GHz amplitude modulation thanks to the proposed all-optical modulation technique, which boosts its achievable phase contrast as compared to a conventional MHz ToF system. Fig. 14 experimentally demonstrates the robustness and depth resolution of our method ascompared to a MHz AMCW ToF camera (LUCID Helios Flex) at the same spatial resolution of 100×100 with the zero-order resampling.

Our nascent method demonstrates overall better recovery of geometric details and comparable shape outlines to the established and commercialized MHz correlation ToF camera. Partly thanks to its light efficiency, our method also generalizes well to challenging materials such as the styrofoam head and the translucent horse. In

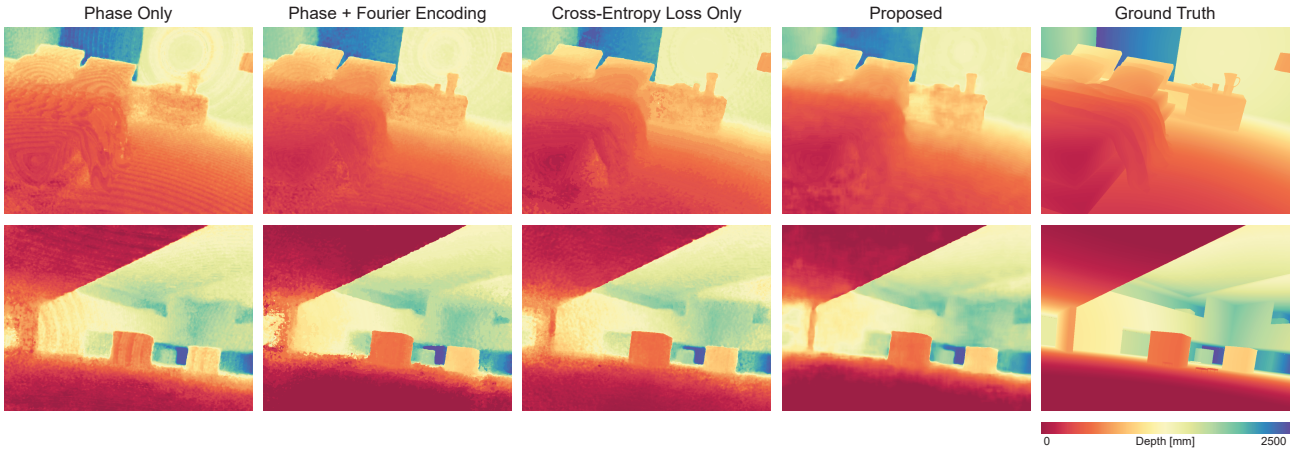

Fig. 9. Reconstruction results for network ablations as defined in Tab. 4. We note successive improvements as we approach the proposed configuration.

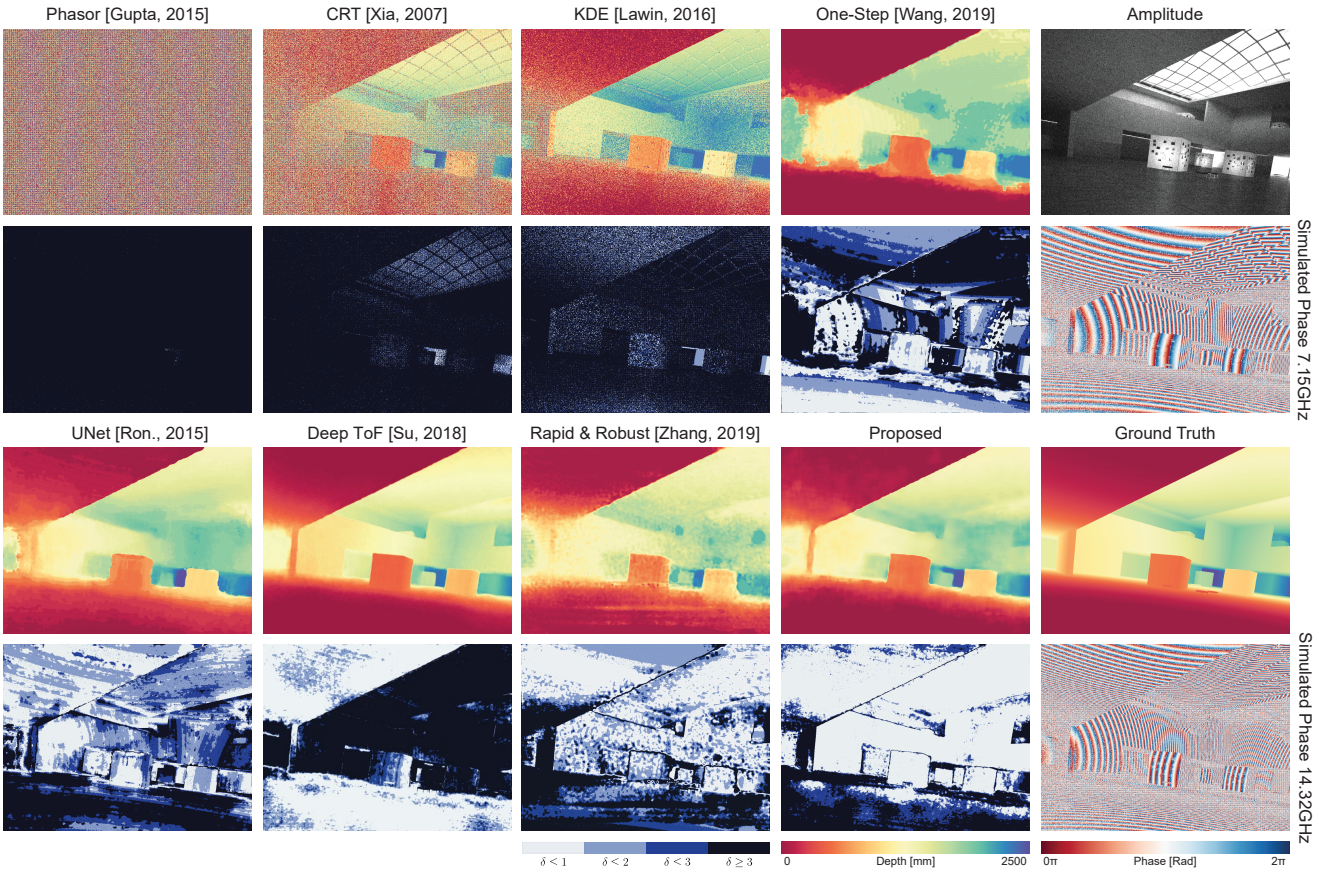

Fig. 10. Additional synthetic results for *box* scene.

particular, light scattering leads to the MHz correlation ToF blending the background and foreground depths of the translucent horse figurine.

10.3 Longer Standoff Distances

In addition to the planar measurement from the main document, we also acquire a portion of a rubber sphere at a longer standoff

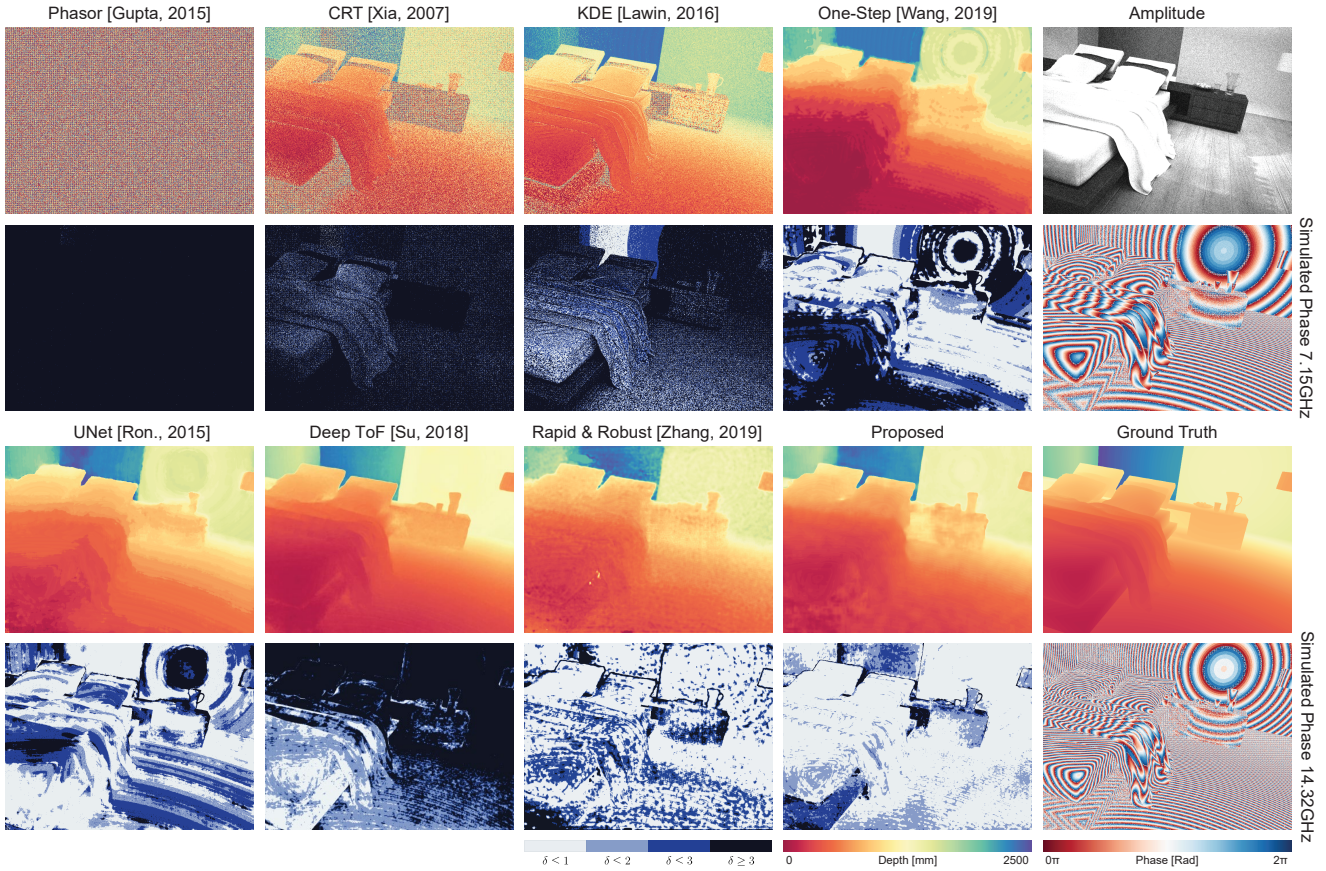

Fig. 11. Additional synthetic results for *bedroom* scene.

distance of 6m (the maximum possible distance for our benchtop setup). The measured portion of the sphere we acquired accurately fits an analytical model of a spherical segment. Fig. 15 shows the experimental reconstruction and the fitted spherical segment.

10.4 Additional Results

In this section we highlight the reconstruction ability of our proposed method on a macroscopic scene with translucent appearance. Fig. 16 shows that even an object comprised of frosted glass, that typical amplitude-modulated continuous-wave time-of-flight systems cannot image, can be faithfully reconstructed with our proposed GHz-frequency all-optical modulation system.

REFERENCES

José Bioucas-Dias, Vladimir Katkovnik, Jaakko Astola, and Karen Egiazarian. 2009. Multi-frequency phase unwrapping from noisy data: adaptive local maximum likelihood approach. In *Scandinavian Conference on Image Analysis*. Springer, 310–320.

David Droeschel, Dirk Holz, and Sven Behnke. 2010. Multi-frequency phase unwrapping for time-of-flight cameras. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1463–1469.

Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd E. Zickler. 2015. Micron-scale Light Transport Decomposition Using Interferometry. *ACM Transactions on Graphics (TOG)* 34, 4 (July 2015).

Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. 2015. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics (ToG)* 34, 5 (2015), 1–18.

Miguel Arevallilo Herráez, David R Burton, Michael J Lalor, and Munther A Gdeisat. 2002. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Applied optics* 41, 35 (2002), 7437–7444.

Felix Järemo Lawin, Per-Erik Forssén, and Hannes Ovrén. 2016. Efficient multi-frequency phase unwrapping using kernel density estimation. In *European Conference on Computer Vision*. Springer, 170–185.

Dingyi Pei, Arto Salomaa, and Cunsheng Ding. 1996. *Chinese remainder theorem: applications in computing, coding, cryptography*. World Scientific.

Rudra PK Poudel, Stephan Liwicki, and Roberto Cipolla. 2019. Fast-scnn: Fast semantic segmentation network. *arXiv preprint arXiv:1902.04502* (2019).

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. 2018. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6383–6392.

Kaiqiang Wang, Ying Li, Qian Kemao, Jianglei Di, and Jianlin Zhao. 2019. One-step robust deep learning phase unwrapping. *Optics express* 27, 10 (2019), 15100–15115.

Xiang-Gen Xia and Genyuan Wang. 2007. Phase unwrapping and a robust Chinese remainder theorem. *IEEE Signal Processing Letters* 14, 4 (2007), 247–250.

Teng Zhang, Shaowei Jiang, Zixin Zhao, Krishna Dixit, Xiaofei Zhou, Jia Hou, Yongbing Zhang, and Chenggang Yan. 2019. Rapid and robust two-dimensional phase unwrapping via deep learning. *Optics express* 27, 16 (2019), 23173–23185.

Zhengyou Zhang. 2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* 22, 11 (2000), 1330–1334.

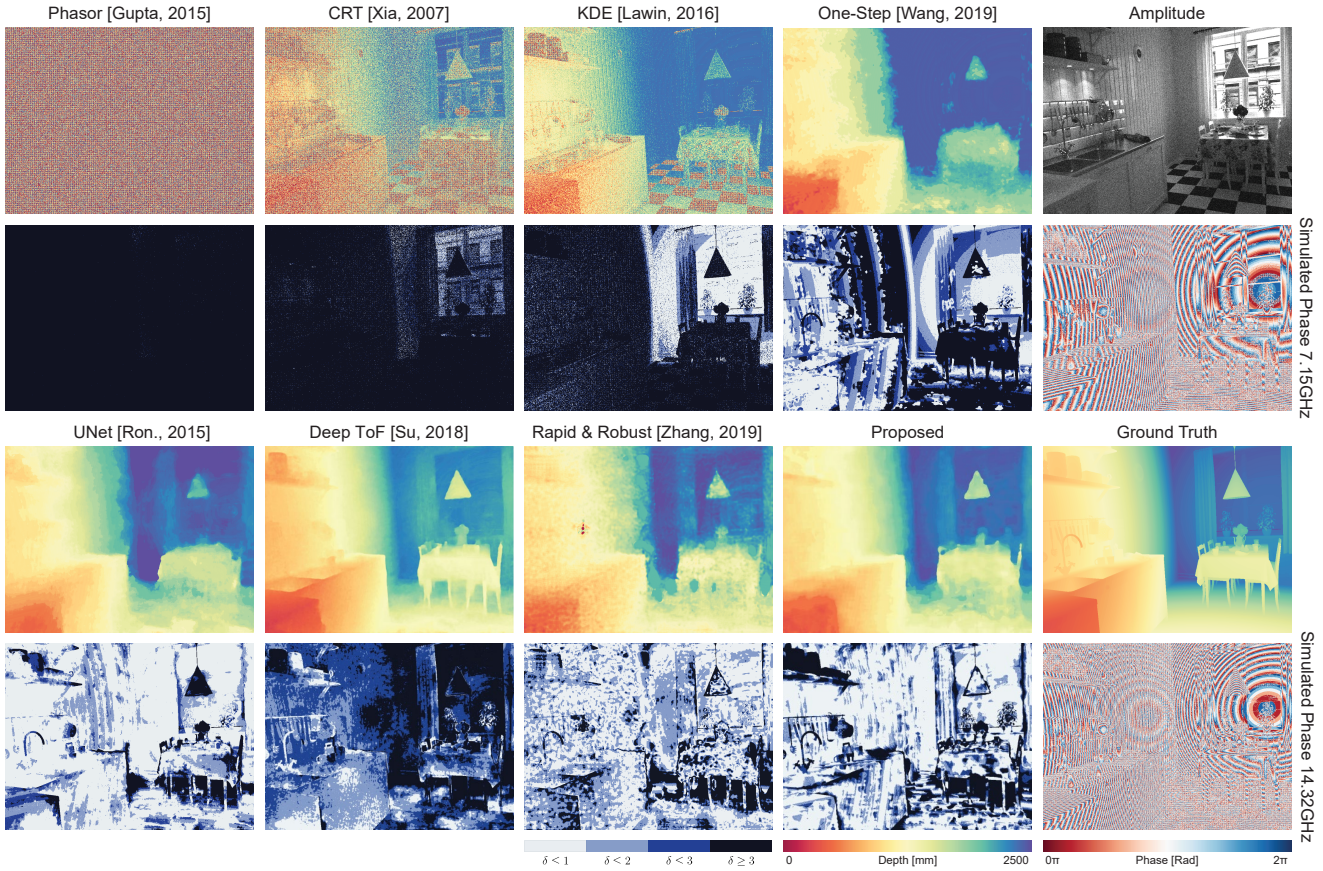

Fig. 12. Additional synthetic results for *kitchen* scene.

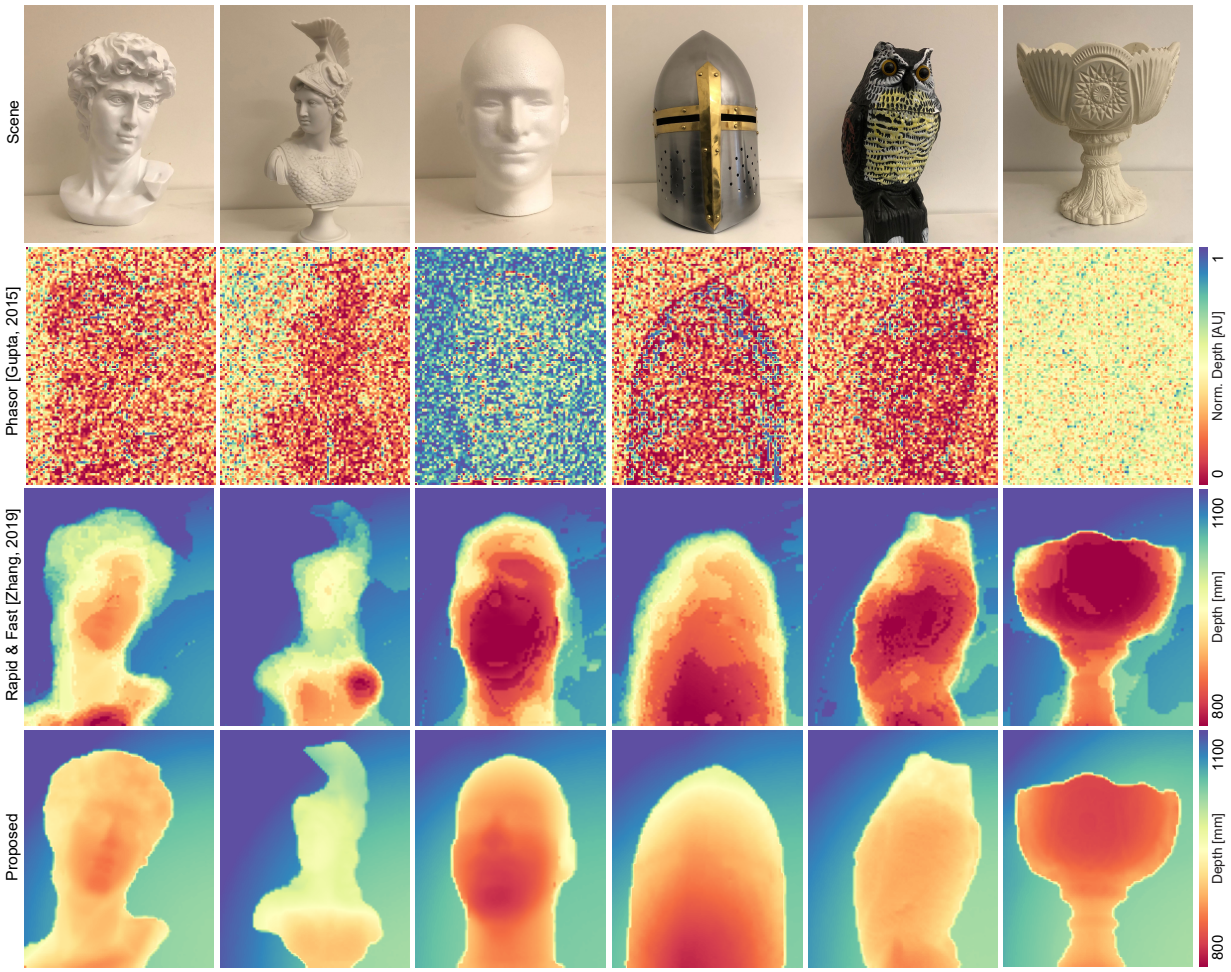

Fig. 13. Our neural unwrapping method outperforms the state-of-the-art learning-based method [Zhang et al. 2019] as well as the established phasor unwrapping method [Gupta et al. 2015] on challenging real data.

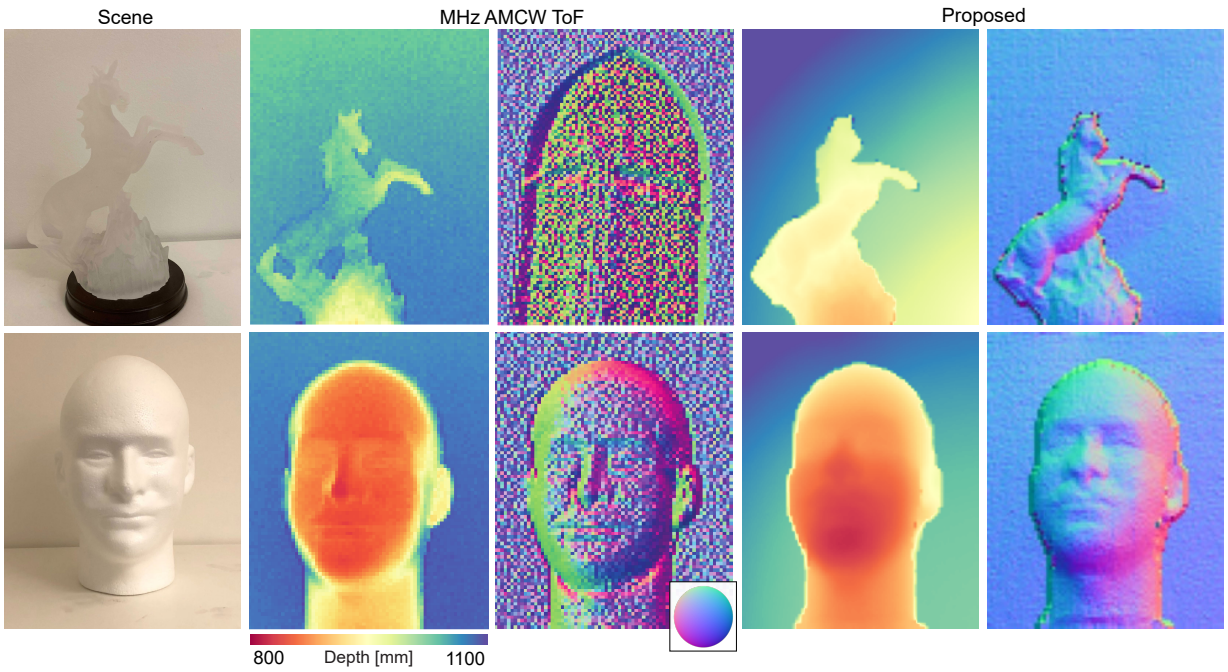

Fig. 14. Our all-optical GHz ToF camera captures fine geometric details even on scenes with challenging reflectance with specularity and wide dynamic range. Specifically, our prototype recovers more consistent surface normals than the MHz ToF, and avoids artifacts such as blending the translucent horse statue with the background or warping the sides of the foam head.

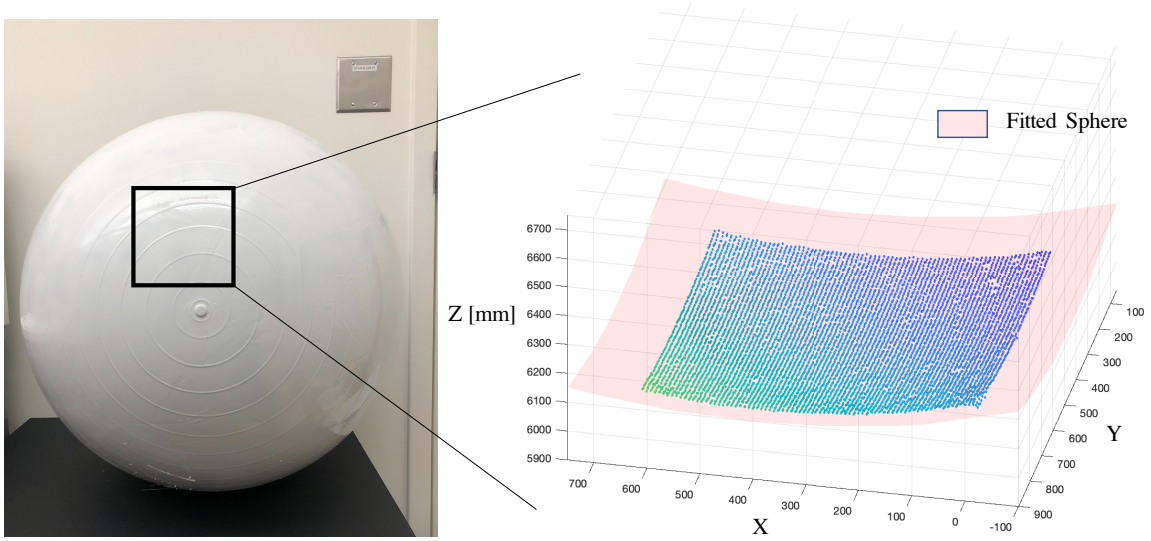

Fig. 15. We acquire a spherical object at 6 m standoff distance that. The recovered measurement accurately fits the spherical geometry.

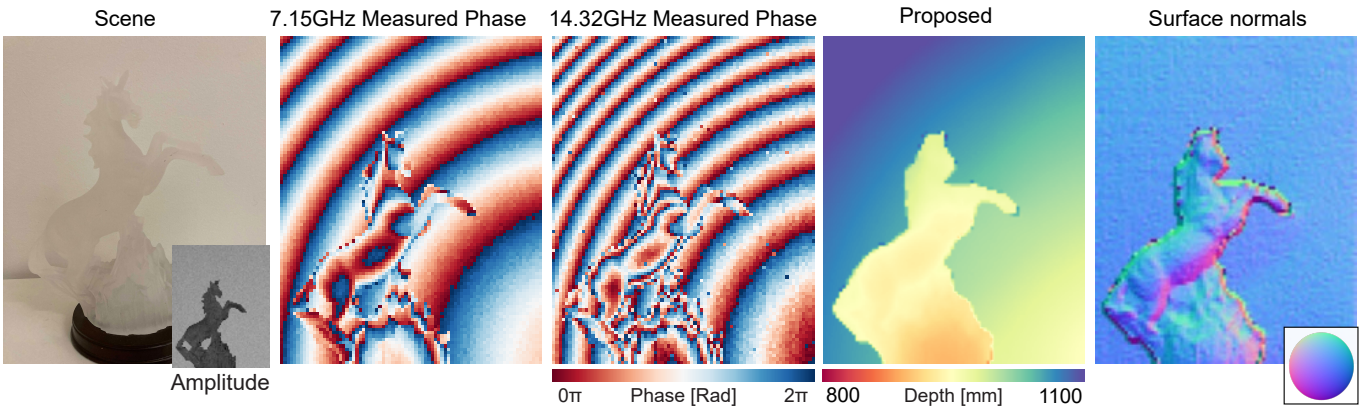

Fig. 16. Results for additional captured scenes with complex surface geometry and/or reflectance properties. Note how te correlation measurements faithfully encode object surface variation.

	Input			Loss		Performance (%)			
	$\hat{\phi}$	$\gamma(\hat{\phi})$	$\lvert\nabla\hat{\phi}\rvert$	$\mathcal{L}_{CE}$	$\mathcal{L}_{L1}$	$\uparrow \delta < 1$	$\uparrow \delta < 2$	$\uparrow \delta < 3$	$\downarrow \delta \geq 3$
Proposed	✓	✓	✓	✓	✓	**51.6%**	**69.1%**	**77.0%**	**23.0%**
$\mathcal{L}_{CE}$ Only	✓	✓	✓	✓	-	49.1%	65.0%	73.3%	26.7%
F. Features	✓	✓	-	✓	✓	40.2%	59.6%	68.8%	31.2%
Phase Only	✓	-	-	✓	✓	30.3%	52.3%	65.7%	34.3%

Table 4. Ablation study configurations and corresponding quantitative results, duplicated from main text for convenience. Here the δ metric represents the percent of pixels whose prediction is δ wraps from ground truth wrap count. Up arrow denotes "higher is better", down arrow means "lower is better".

Method	$\uparrow \delta < 1$	$\uparrow \delta < 2$	$\uparrow \delta < 3$	$\downarrow \delta \geq 3$	$\downarrow \delta \geq 10$
Phasor [2015]	0.74%	1.66%	3.50%	96.5%	84.4%
CRT [2007]	9.29%	14.7%	19.7%	80.3%	56.0%
KDE [2016]	9.46%	18.56%	27.0%	73.0%	8.93%
One-Step [2019]	19.9%	37.6%	52.2%	47.8%	14.6%
U-Net [2015]	21.8%	45.6%	64.4%	35.6%	10.0%
Deep-ToF [2018]	20.1%	47.5%	67.6%	32.4%	8.4%
Rapid. [2019]	23.1%	45.4%	61.1%	38.9	9.74%
Proposed	**51.6%**	**69.1%**	**77.0%**	**23.0%**	**7.6%**

Table 5. Quantitative comparison table for proposed neural phase unwrapping method and baselines, as evaluated on the synthetic test scenes. Repeated from main text for convenience.