# WaveMo: Learning Wavefront Modulations to See Through Scattering

## Supplementary Material

In this document, we provide additional experimental results and ablation studies. Section 1, 2 and 3 shows additional results on static in-distribution targets, static out-of-distribution targets, and dynamic targets, respectively. Section 4 are ablation studies regarding the number and the type of modulations. Section 5 reports the standard deviation of quantitative performance on real data.

Videos of our dynamic results are shown in our project webpage at https://wavemo-2024.github.io/.

## 1. Static In-distribution Scenes

Supplementary to Figure 5 in the main paper, this section provides additional evaluation results on target scenes that belong to the same type of human body tissue (adipose) as those used for training. Here, we use the proxy network for reconstruction. As can be seen from the red zoom-in boxes in Figure 9, reconstructions with learned modulations show significantly better performance.

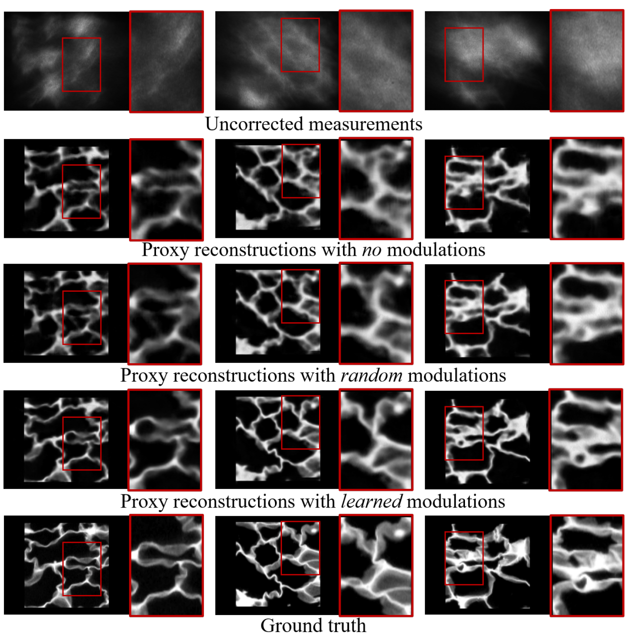


Figure 9. **Proxy Network Reconstruction of In-distribution Targets (supplementary to Figure 5 in the main paper).** Zoomed-in regions are labeled in red boxes. Reconstructions using learned modulations contain much finer details.

## 2. Static Out-of-Distribution Scenes

This section provides additional evaluation results on human pathological tissue slides that do not appear in the training set, including sections of parotid, stomach, and finger. Section 2.1 shows results using the proxy reconstruction network, while Section 2.2 shows results using an unsupervised iterative approach [8].

### 2.1. Proxy Reconstruction Network

Results using a proxy reconstruction network are demonstrated in Figure 10 and Table 4. Reconstructions with learned modulates outperform those with random modulations and those without modulations.

| Metric | Modulations | | |
|---|---|---|---|
| | None | Random | Learned |
| PSNR (SD) | 16.69 (0.37) | 17.17 (0.36) | **18.05 (0.32)** |
| SSIM (SD) | 0.47 (0.026) | 0.49 (0.024) | **0.56 (0.019)** |

Table 4. **Results on Out-of-distribution Scenes Using a Proxy Network Equipped with Learned Modulations.** The metrics are averaged over 100 samples. We also report the standard deviation (SD) for both PSNR and SSIM. Our learned modulations achieve the best performance.

### 2.2. Unsupervised Iterative Approach

Results using an unsupervised iterative approach [8] are demonstrated in Figure 11 and Table 5. Similar to our previous observations, reconstructions with the 16 learned modulations exhibit clearer shapes and enhanced contrast than those with random or no modulations.

| Metric | Modulations | | |
|---|---|---|---|
| | None | Random | Learned |
| PSNR (SD) | N/A | 10.64 (0.78) | **12.73 (0.42)** |
| SSIM (SD) | N/A | 0.23 (0.035) | **0.32 (0.028)** |

Table 5. **Results on Out-of-distribution Scenes Using an Unsupervised Iterative Approach Equipped with Learned Modulations.** The metrics are averaged over six samples. We also report the standard deviation (SD) for both PSNR and SSIM. The unsupervised iterative method [8] relies on multiple wavefront modulations and therefore cannot recover objects with a single measurement (no modulation), hence the "N/A". Compared with random modulations or no modulations, learned modulations lead to better reconstructions.

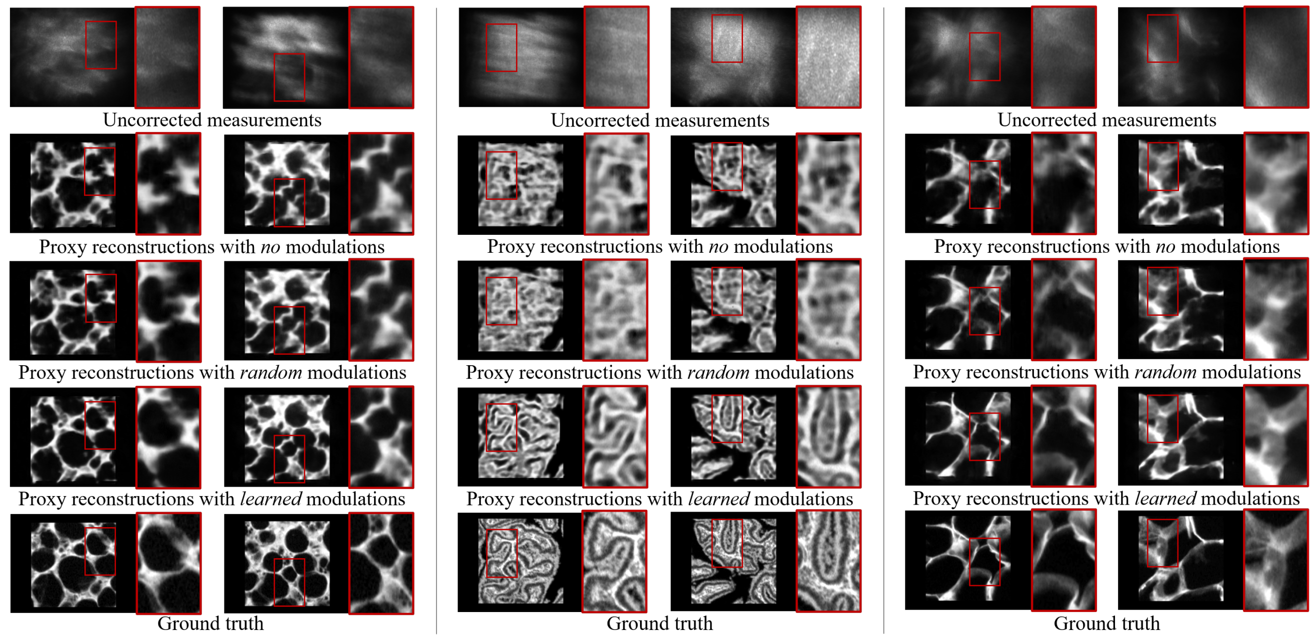


Figure 10. **Results on Out-of-distribution Scenes using a Proxy Network Equipped with Learned Modulations.** From left to right, the 3 columns show human pathological tissue sections of parotid, stomach, and finger, respectively. Zoomed-in regions are labeled with red boxes. Learned modulations significantly enhance the reconstruction quality.

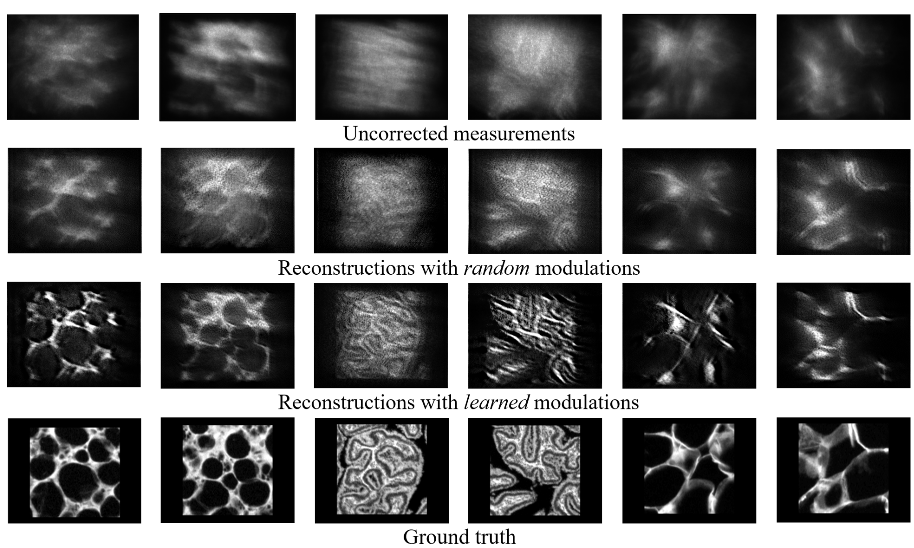


Figure 11. **Results on Out-of-distribution Scenes Using an Unsupervised Iterative Approach [8] Equipped with Learned Modulations.** The target scenes include the parotid (left two columns, stomach (middle two columns), and finger (right two columns). Reconstructions with learned modulations achieve the best quality.

## 3. Dynamic Out-of-Distribution Scenes

Supplementary to Figure 8 in the main paper, we provide two additional sets of dynamic experiments using the unsupervised iterative approach [8]. Same as the experiments in Figure 8, each dynamic scene in Figure 12 contains 48 frames, which are captured by cycling our learned 16 modulations. In the first scene, the two digits 1 and 2 are translated in opposite directions. In the second scene, a flower undergoes a counterclockwise rotation of 0.5° per frame. Even though the results obtained using learned modulations still suffer from artifacts, they are still much

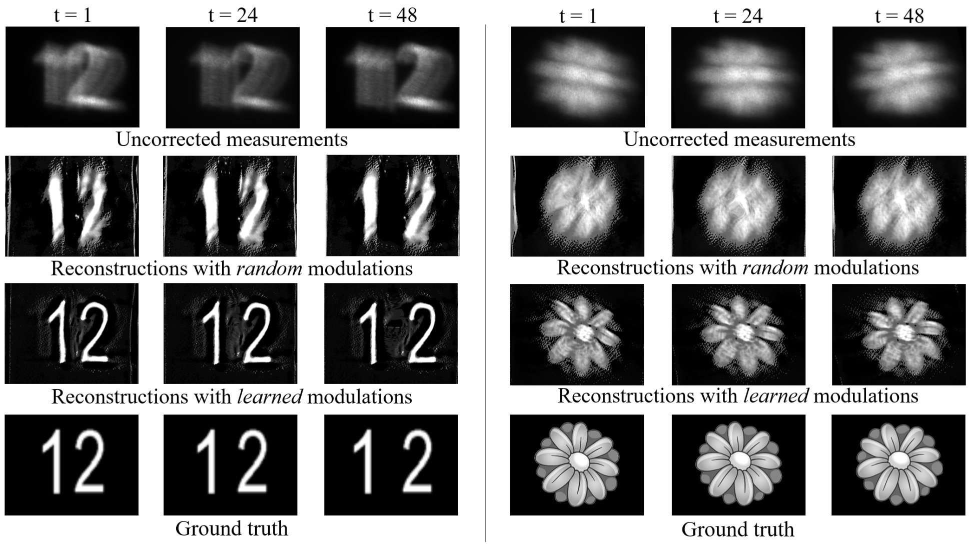


Figure 12. **Unsupervised Reconstruction of Two Additional Dynamic Scenes. (Videos are shown in the bottom of the webpage provided in the supplementary Zip file).** In the left scene, two digits are moving in opposite directions; in the right scene, a flower is rotating. Reconstructions are done using an unsupervised iterative approach [8]. Our learned modulations achieve the best reconstructions.

sharper than results using random or no modulations.

# 4. Ablation Studies

## 4.1. Number of Modulations

We analyze the impact of the number of modulations in simulation. As shown in Table 6, PSNRs of both random (sampled from Zernike space) and learned modulations exhibit a similar trend of improvement as $K$ increases. Note that 4 learned modulations beat 32 random modulations by over 2 dB.

| Modulation | $K = 4$ | $K = 8$ | $K = 16$ | $K = 32$ |
|---|---|---|---|---|
| Random | 25.550 | 25.962 | 26.439 | 26.650 |
| Learned | 28.996 | 29.916 | 30.391 | 30.550 |

Table 6. **Quality (PSNR) v.s. Number of Modulations ($K$).** PSNRs of both random (sampled from Zernike space) and learned modulations exhibit a similar trend of improvement as $K$ increases. Note that 4 learned modulations beat 32 random modulations by over 2 dB.

## 4.2. Different Types of Modulations

We compare our learned approach against several heuristic approaches for the design of modulation patterns, such as per-pixel random Gaussian matrices, random Zernike polynomials, focus sweeping, and directly optimizing the MTF. This comparison is done in simulation. As shown in Table 7, our learned modulations notably outperform these heuristic designs.

| PSNR | Gauss | Zern | Focus | MTF | Ours |
|---|---|---|---|---|---|
| Mean | 26.322 | 26.439 | 26.096 | 25.917 | **30.391** |
| SD | 0.38 | 0.16 | 0.26 | 0.12 | 0.22 |

Table 7. **Quality (PSNR) v.s. Modulation Types.** We compare our learned approach against per-pixel random Gaussian matrices (Gauss), random Zernike polynomials (Zern), focus sweeping (Focus), and directly optimizing the MTF. Our approach outperforms the baselines by over 3 dB.

# 5. Standard Deviation of Results on Real Data

To supplement the quantitative evaluation on real data in Tables 2 and 3, we report the Standard Deviation (SD) as a further statistical measure, shown in Tables 8 and 9.

| Method | Data type | Modulations | | |
|---|---|---|---|---|
| | | None | Random | Learned |
| Proxy | Tissue | 16.53 (0.34) | 17.57 (0.32) | **19.06 (0.29)** |
| Proxy | Out-of-dist. | 9.34 (0.38) | 9.90 (0.37) | **10.71 (0.35)** |
| Iterative [8] | Static | N/A | 11.26 (0.78) | **14.61 (0.45)** |
| Iterative [8] | Dynamic | N/A | 8.90 (0.51) | **12.89 (0.33)** |

Table 8. **PSNR and Standard Deviation of Experimental Results.** For our jointly trained feed-forward proxy reconstruction network ("proxy"), we tested on 40 tissue samples and 8 out-of-distribution scenes, all of which are static. For the iterative method [8], we tested on the same 8 out-of-distribution static scenes. We also tested [8] on 2 dynamic scenes, each with 48 frames. The iterative method relies on multiple wavefront modulations and therefore cannot recover objects with a single measurement, hence the "N/A". Standard deviations (SD) are included in the parentheses. Compared against random modulations or no modulations, our learned modulations lead to better reconstruction performance for both the proxy network and an unsupervised iterative approach.

| Method | Data type | Modulations | | |
|---|---|---|---|---|
| | | None | Random | Learned |
| Proxy | Tissue | 0.44 (0.023) | 0.48 (0.021) | **0.58 (0.016)** |
| Proxy | Out-of-dist. | 0.29 (0.029) | 0.30 (0.028) | **0.32 (0.028)** |
| Iterative [8] | Static | N/A | 0.21 (0.034) | **0.38 (0.031)** |
| Iterative [8] | Dynamic | N/A | 0.23 (0.032) | **0.33 (0.029)** |

Table 9. **SSIM (SD) of Experimental Results.** Same as Table 8 but with SSIM. Standard deviations (SD) are included in the parentheses. Our learned modulations lead to better performance.