

WIRE: Wavelet Implicit Neural Representations (Supplementary Document)

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan,
Ashok Veeraraghavan, Richard G. Baraniuk
Rice University

<https://vishwa91.github.io/wire>

1. Document Overview

This document supplements the main paper with additional details about implementation, and further insights into WIRE and its effect on signal representation accuracy, effect of initialization, visualization of layer outputs, and performance for various inverse problems.

2. Experiments

2.1. Choice of wavelet function

We showed that a complex Gabor wavelet activation function of the form $e^{j\omega_0 x} e^{-|s_0 x|^2}$ enabled high capacity and robust representation for visual signals. Here, we show that other continuous wavelets can also be used to equip implicit neural representations (INRs) with performance comparable to that of complex Gabor wavelet. We compared complex Gabor wavelet, real Gabor wavelet ($\cos(\omega_0 x) e^{-(s_0 x)^2}$), mexican hat wavelet ($(1 - (s_0 x)^2) e^{-(s_0 x)^2}$), and difference of Gaussian ($e^{-(s_0 x)^2} - e^{-(s_1 x)^2}$). Figure 1 shows two examples on image representation. In first example (a), we show representation accuracy for a noise-free settings. In (b), we show representation accuracy for a noisy image, and (c) shows visualization of final output for noisy image representation. Across the board, we see that the complex Gabor wavelet has superior performance, both in terms of speed of representation, as well as robustness, which motivated our choice of wavelet in the main paper.

2.2. WIRE initialization

INRs like SIREN [5] strongly depend on initialization to obtain accurate representation. WIRE does not require any initialization except for the default uniform weights. However, since WIRE consists of a complex sinusoidal term, it marginally benefits from SIREN-like initialization. To understand the dependence, we evaluated approximation accuracy for image representation (no noise), and image denoising (20dB image noise). Here, a SIREN-like weight initialization implies the first layer weights are drawn from

$\mathcal{U}(-1/N, 1/N)$ and the weights of the rest of the layers are drawn from $\mathcal{U}(-\sqrt{6}/(\omega_0 N), \sqrt{6}/(\omega_0 N))$, where N is the number of input features and $\mathcal{U}(a, b)$ is a uniform distribution over $[a, b]$. A normal weight initialization involves drawing weights from $\mathcal{U}(-1/\sqrt{N}, 1/\sqrt{N})$ for all layers. Fig. 2 compares the representation accuracy for SIREN-like and standard initialization. In both cases, we see that the trends are nearly similar; SIREN-like initialization results in up to 1dB higher accuracy. Hence WIRE is largely robust to initial parameters which enables easy tuning across a large range of hyperparameters ω_0, s_0 .

2.3. WIRE layer visualizations

Gabor wavelets uniquely enable space–frequency localization, a property we observe is inherited by WIRE. To evaluate this hypothesis, we visualized the output WIRE composed of an MLP with two hidden layers and 181 hidden features each. We then learned a representation for a Siemens star test image that consists of all spatial frequencies and orientations. Fig. 3 visualizes the input image real and imaginary outputs of 64 hidden features with least variance. The output of each first layer feature consists of one-dimensional Gabor wavelets at various orientations, while the outputs of second layer consist of sparsely populated images.

Fig. 4 visualizes outputs at each layer for various nonlinearities and the final approximated image for the Siemens star test image. The sparse outputs of second layer are evidently unique to WIRE. Gauss has outputs that look less spares, while SIREN and ReLU with positional encoding result in dense outputs. This has a direct consequence on approximation capacity for high frequency parts of the signal. The final result in the bottom row shows that the sparse nature of outputs of WIRE enables high approximation accuracy with qualitatively better features at the center of the image which consists of highest spatial frequencies. Gauss follows next as it results in the second most sparse outputs at each layer. SIREN and ReLU with positional encoding alike produce blurry outputs at the center, primarily due to the non-compact nature of outputs. WIRE’s ability to de-

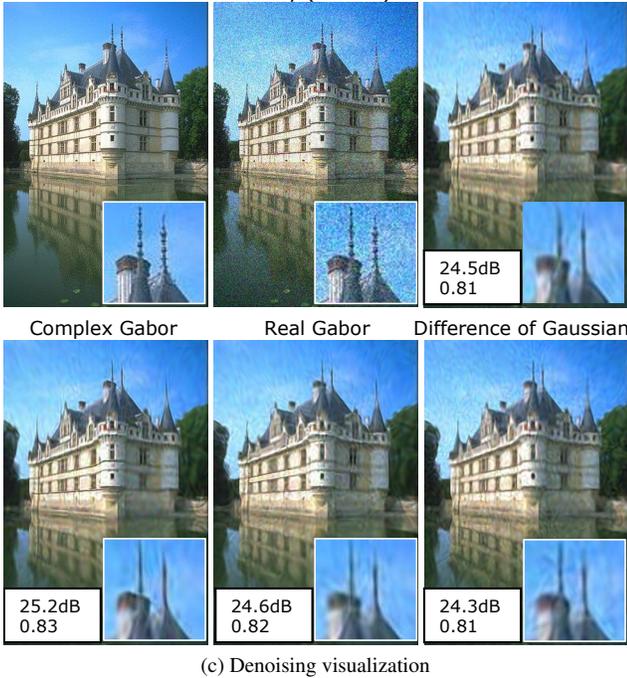
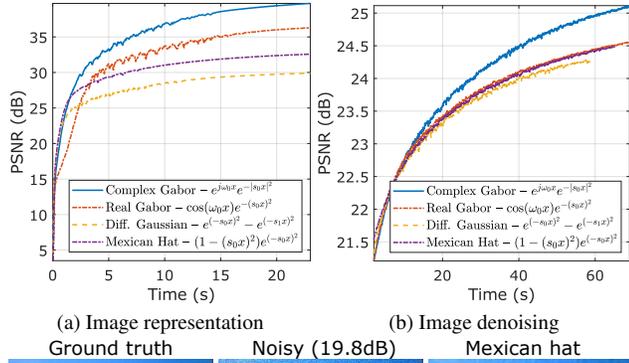


Figure 1. **Effect of wavelet.** WIRE can be equipped with various forms of nonlinearities including a complex gabor wavelet (as shown in the main paper), real gabof wavelet, mexican hat, and difference of Gaussians. The figure above shows (a) accuracy vs. time for image representation, (b) accuracy vs. time for image denoising, and (c) visualization of denoised results. While the results are comparable, complex Gabor results in highest accuracy.

compose images as a linear combination of sparse images results in high representational capacity for the same number of parameters, as we verify empirically in the next section.

2.4. Sensitivity to training parameters

WIRE is a promising INR model that achieves high representation accuracy and is robust to a wide range of training parameters. We demonstrate the efficacy of WIRE in this section with several sensitivity analyses.

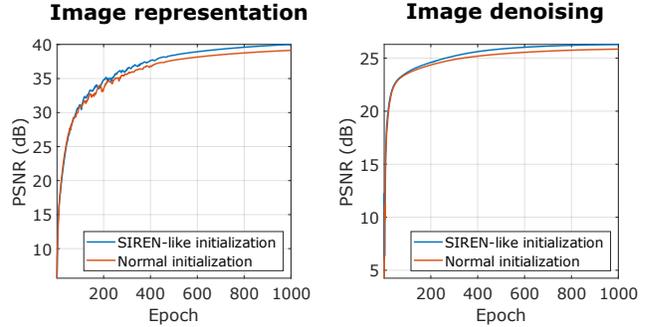


Figure 2. **Effect of initialization.** The plots show approximation accuracy for image representation and image denoising (20dB input PSNR) across training epochs with SIREN-like weight initialization and standard initialization. WIRE is robust to the initial weights, but marginally benefits from a SIREN-like initialization.

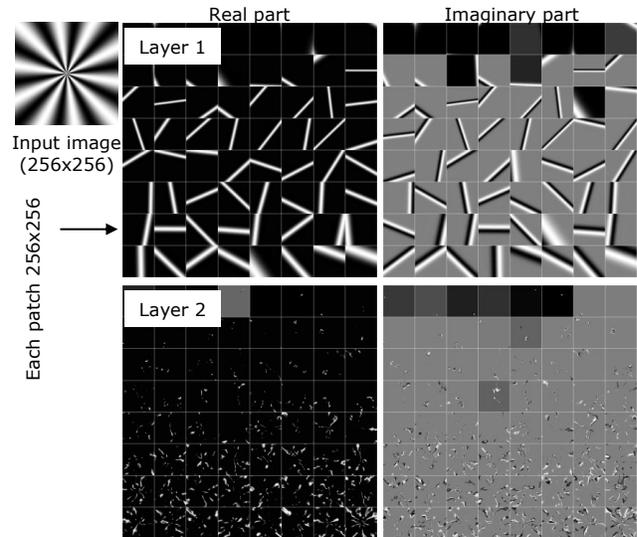


Figure 3. **Layer outputs for WIRE.** The input image is a Siemens star test image that contains all spatial frequencies and all angles. The patches show outputs (same size as image) of each hidden feature in layer one and two for a two hidden-layer MLP equipped with WIRE nonlinearity. WIRE results in sparse images which enables high representational capacity for images, as shown in Fig. 4.

Effect of learning rate. WIRE performs well for a large range of learning rates. To understand the performance trends, we learned an image representation with added noise (20dB input PSNR) with various nonlinearities. We used a 2-layer MLP with 256 hidden features per layer. Fig. 5 shows the maximum representation PSNR with varying learning rates. WIRE has a stable and significantly higher accuracy compared to other approaches. Interestingly, the highest accuracy is achieved at a high learning rate of 2×10^{-2} . This behavior is also observed with deep image prior [7] where a larger learning rate enabled stronger regularization. Such a similar behavior implies WIRE en-

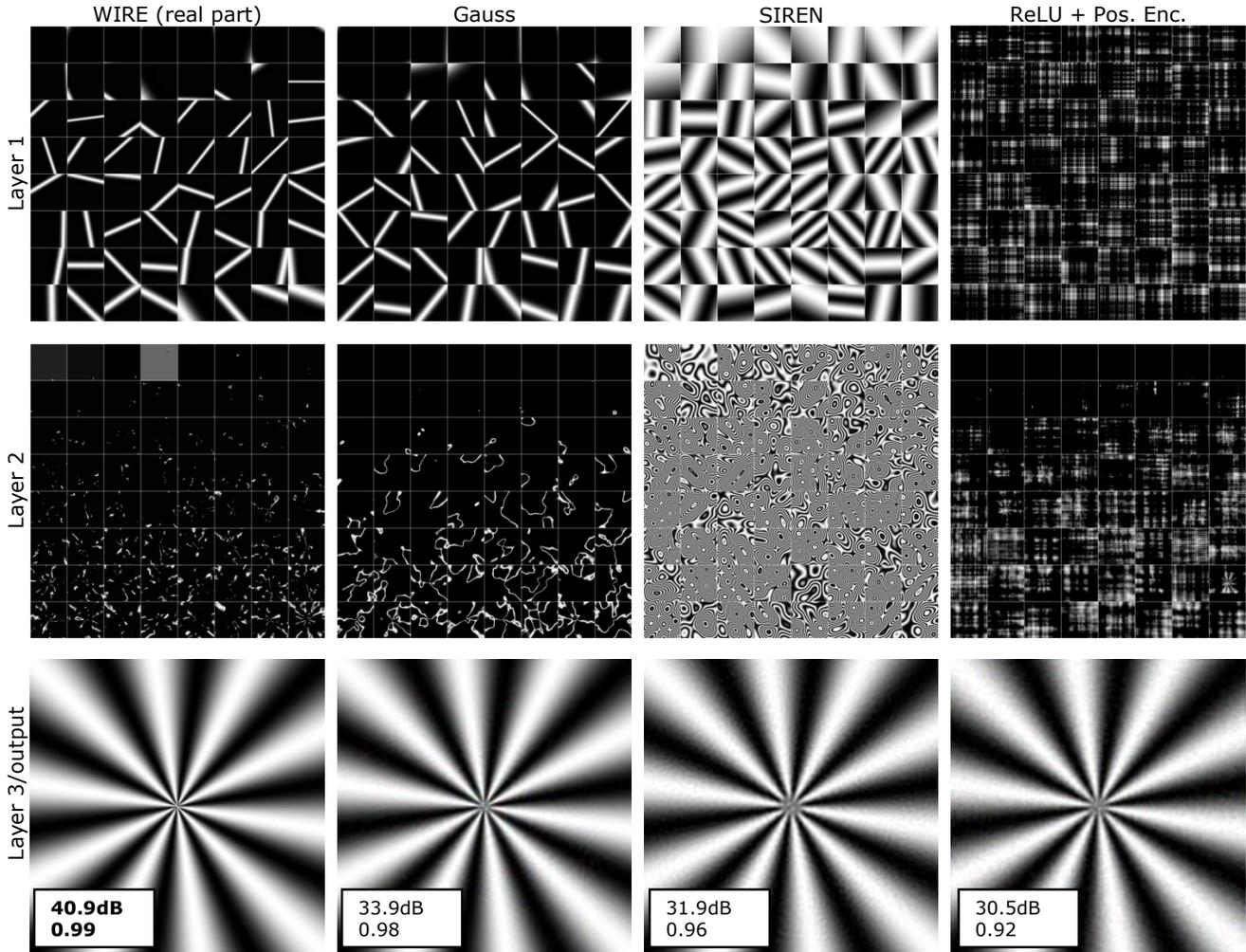


Figure 4. **Visualization of hidden layer outputs.** The figure above visualizes outputs of hidden features in the two layers for the Siemens sector test image shown in Fig. 3. WIRE uniquely results in sparse images, which enables high accurate representation of high frequency parts of the image (center of the sector).

joys strong inductive biases and hence is amenable to solve inverse problems.

Effect of number of layers. Fig. 6a shows a plot of representation accuracy of an image for varying number of hidden layers with various nonlinearities. In each case, the number of hidden features were set to 256. We reduced the learning rate with increasing layers to avoid divergence. WIRE uniformly outperforms other approaches (except with 0 hidden features), as is to be expected as Gabor wavelets enable high approximation accuracies for images. Interestingly, for a large number of hidden layers (≥ 3), WIRE performance is similar to SIREN and Gaussian non-linearity. This is to be expected as the network has a large capacity with so many layers. However, a large number of layers is computationally expensive and often results in

an unstable learning regime. WIRE therefore is a reliable choice for small to medium number of hidden layers for most cases.

Effect of number of features. Fig. 6b shows approximation accuracy for image representation with varying number of hidden features. In all cases, the number of hidden layers were fixed to be two. The performance of WIRE is similar to other nonlinearities at very low number of hidden features, where all models similarly lack sufficient richness. For higher than 128 features, WIRE outperforms other approaches with MFN [4] coming a close second.

Limitations of WIRE. For signals that have non-compact structure, such as periodic signals, or chirp-like images, WIRE performs poorer than approaches such as SIREN.

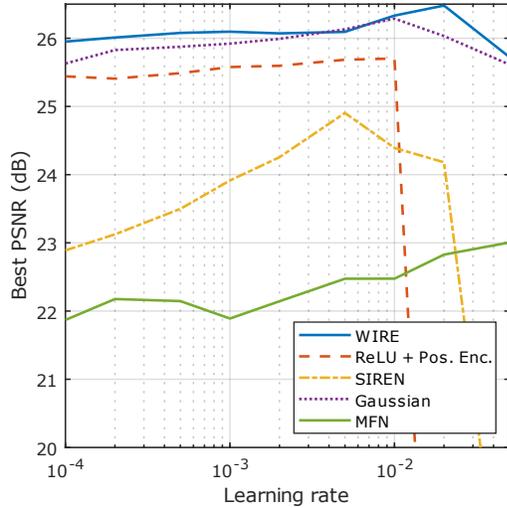


Figure 5. **Effect of learning rate.** The plot above shows approximation accuracy for representing a noisy image (input PSNR of 20dB) with various nonlinearities. WIRE is robust to learning rate, and produces best results with high learning rate of 2×10^{-2} .

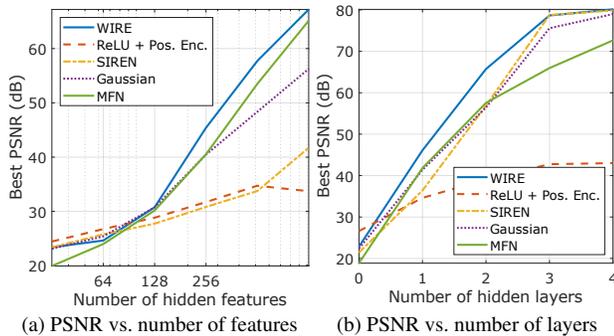


Figure 6. **Effect of number of parameters.** The plot above shows approximation accuracy for representing an image with varying number of (a) hidden features and (b) hidden layers. WIRE outperforms other nonlinearities with 128 or more hidden features, and one or more layers and is nearly the same as SIREN and Gaussian nonlinearities for more than 3 layers.

Figure 7 visualizes an example of representing a 2D chirp with SIREN And WIRE. Evidently, SIREN results in 6dB higher accuracy compared to WIRE. Such structures are rare in natural images, but benefit better from nonlinearities that are non-compact such as SIREN.

2.5. Inverse problems

Computed tomographic reconstruction. We showed in the main paper that computed tomography (CT) benefits from inductive biases of INRs. Here, we study the effect of number of measurements. Fig. 8a shows the ground truth image we used in our experiments. We denoised the original 512×512 image [2] with BM3D [3] ($\sigma = 0.1$) to re-

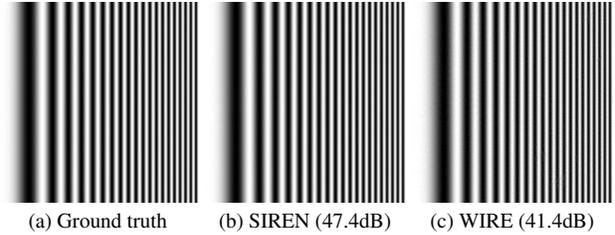


Figure 7. **WIRE limitations.** For signals with non-compact structure such as the 2D ramp shown above, SIREN performs better than WIRE, which is expected as the sinusoidal nonlinearity is non-compact.

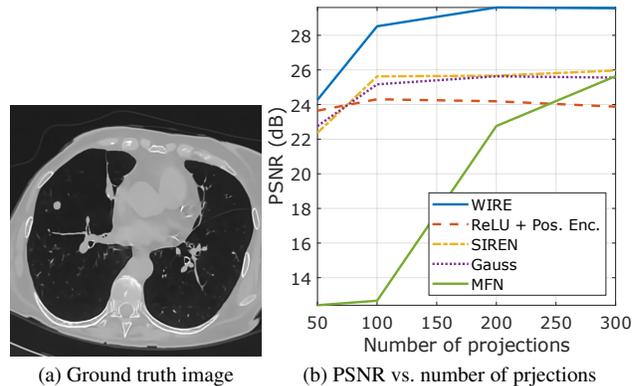


Figure 8. **CT with varying number of projections.** (a) shows the 512×512 ground truth x-ray image of lungs [2] we used for our CT experiments. We denoised the original image to remove streak artifacts. (b) shows accuracy as a function of number of measurements with various nonlinearities. Across the board, WIRE outperforms all other approaches by a considerable margin.

move streak artifacts. We then simulated CT reconstruction with varying numbers of projections. In each case, we used an MLP with three hidden layers and 256 hidden features per layer. We sampled the INR on a regular grid to first generate the image, and then use Radon transform to obtain the sinogram. From the accuracy plot in Fig. 8b, we see that WIRE achieves higher PSNR than any other nonlinearity. Fig. 9 visualizes the reconstruction with varying number of projections for each nonlinearity. The reconstruction is visually superior even with small number of projections, which is particularly beneficial for reducing exposure to x-rays during capture.

Multi-image super-resolution. We showed a result on multi-image super-resolution in the main paper. Here, we provide more details about the experiment. Figure 10 shows the 512×768 dimensional ground truth image from the Kodak dataset [1]. We simulated a total of four low-resolution images by modifying each $4 \times$ downsampled image by a small translation and rotation, thereby resulting in sub-pixel

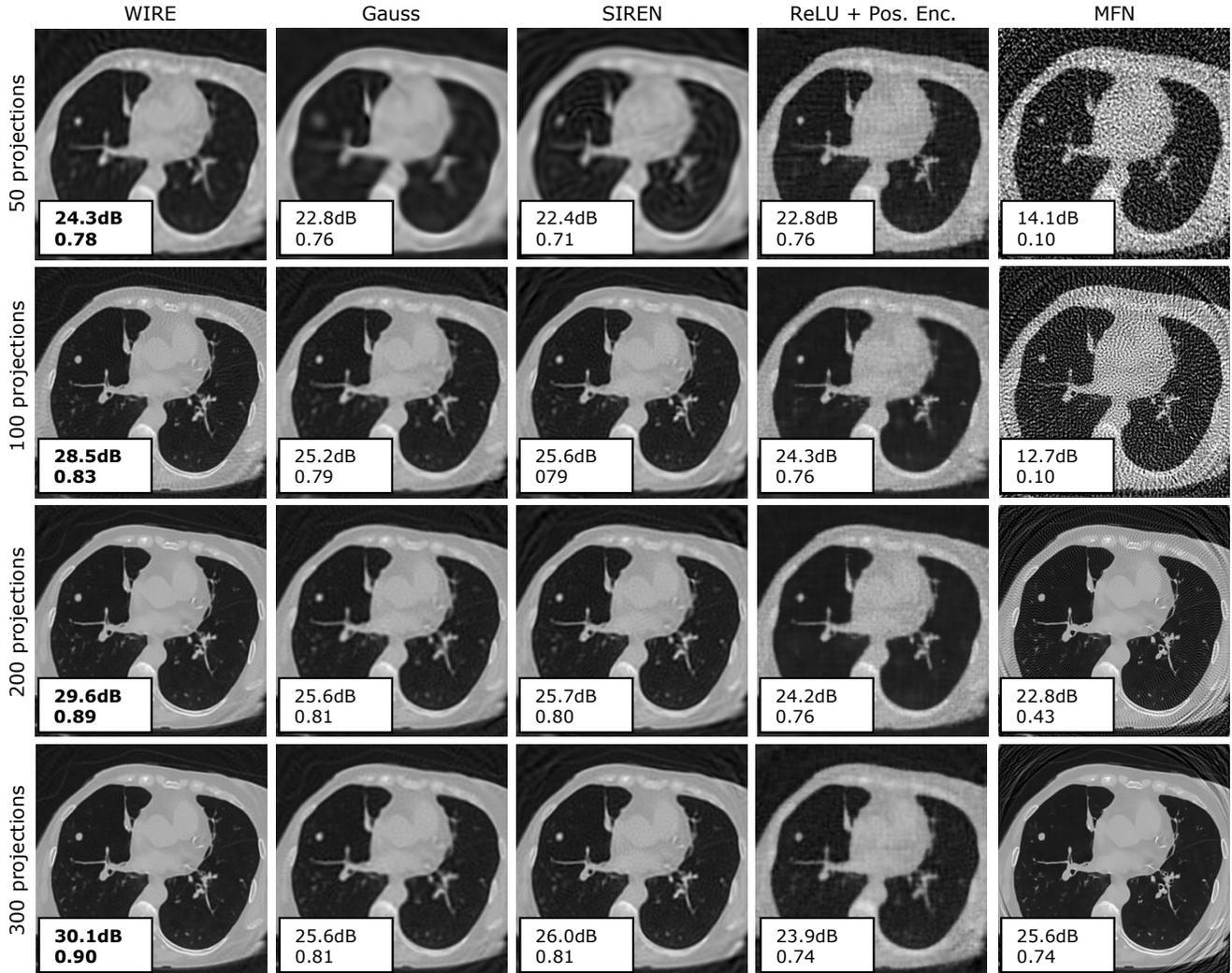


Figure 9. **Effect of number of projections on CT accuracy.** The images above visualize reconstruction for the lungs image shown in Fig. 8a. WIRE outperforms all other approaches even with 50 projections (10% measurements) and is visually pleasing.

motion between the frames. We assumed the transformation A_k between the high-resolution frame \mathbf{x} and each low-resolution frame \mathbf{y}_k was known. We represented the high resolution \mathbf{x} as output of an INR. In each case, the INR had three hidden layers with 256 hidden features. We then solved a linear inverse problem to estimate the high resolution image. Figure 10 shows the reconstructed output for each nonlinearity and their metrics. The inset shows reconstruction of spokes in the motorcycle. Visually, WIRE generates the sharpest features without any ringing artifacts. Moreover, WIRE results in 1dB or better reconstruction accuracy, and 0.04 higher SSIM.

2.6. Neural radiance fields

Implementation details. For experiments both in main paper and the supplementary, we used the `torch-ngp`

package [6] that implements a wide variety of approaches for training neural radiance fields. The architecture consists of two networks that predict transmittance (σ) and the color at each voxel respectively. Each of the two networks consisted of an MLP with four hidden layers and 182 hidden features each. The color MLP took position (x, y, z) and direction (θ, ϕ) as inputs, while the transmittance MLP took only the position as input. As with all other experiments, we used $182/\sqrt{2} = 128$ hidden features for WIRE to account for parameter doubling due to complex weights. We downsampled the images by $4\times$ to ensure that the model and training data fit in the graphical processing unit’s (GPU) memory. In the main paper, we used a total of 25 randomly chosen images to train the NeRF, and then validated it on 100 images. We used a learning rate of 4×10^{-4} for WIRE and 2.5×10^{-4} for all other nonlinearities and reduced it

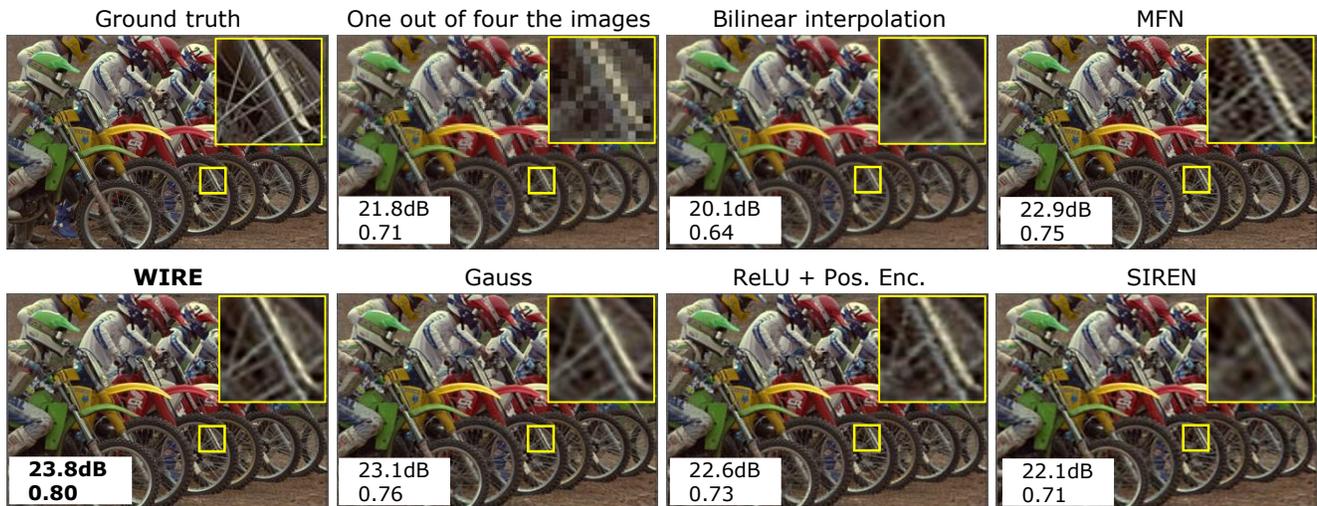


Figure 10. **Multi-image super-resolution.** The figure above visualizes multi-frame super-resolution where each sub-frame was simulated with a small known sub-pixel shift. WIRE achieves highest reconstruction accuracy with qualitatively better reconstruction.

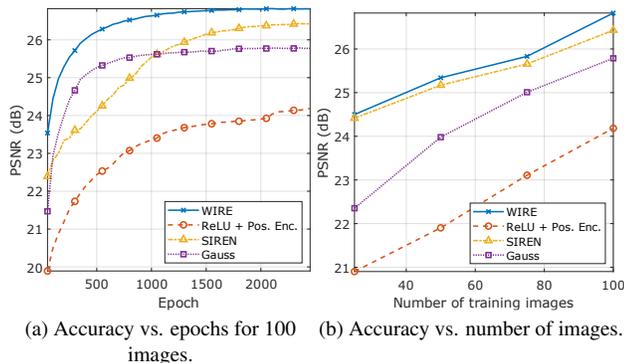


Figure 11. **NeRF accuracy on drums dataset.** (a) shows training accuracy at each epoch for various nonlinearities with neural radiance fields when trained with 100 images. WIRE achieves 0.4dB higher than the next highest (SIREN) when trained with 100 images and does so in a rapid manner. (b) shows accuracy as a function of number of images with WIRE outperforming other approaches for all number of images.

to $0.1\times$ initial value over a total of 2500 training epochs. Except for ReLU, we did not use any form of positional encoding with other nonlinearities as we wished to demonstrate the capacity of each nonlinearity by itself.

Effect of number of images. Fig. 11a shows accuracy vs. number of epochs for the drums dataset when trained with all 100 images. WIRE results in highest accuracy within 2500 epochs and converges more rapidly than other approaches. Fig. 11b shows accuracy as a function of number of training images. WIRE achieves 0.1dB higher than the next competitor SIREN for 25, 50, and 75 images, and

0.4dB higher when trained with 100 images.

Fig. 12 visualizes one of the reconstructed views for the drums. We varied the number of images from 25 to 100 and then rendered the image from a novel view. Visually, WIRE generated the most pleasing results including sharp features of the cymbals and their stands, and the smooth membrane on the drum. In contrast, Gaussian nonlinearity results in cloudy artifacts, while SIREN has high frequency artifacts, especially at lower numbers of images. ReLU+positional encoding requires all 100 images and considerably more than 2500 epochs to reconstruct the components. In all, WIRE is a robust solution for training radiance fields, even with a small number of training samples.

References

- [1] Kodak lossless true color image suite. <http://r0k.us/graphics/kodak/>, 1999. Accessed: 2022-11-09. 4
- [2] Samuel G. Armato III, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao, Denise R. Aberle, Claudia I. Henschke, Eric A. Hoffman, Ella A. Kazerooni, Heber MacMahon, Edwin J. R. van Beek, David Yankelevitz, Alberto M. Biancardi, Peyton H. Bland, Matthew S. Brown, Roger M. Engelmann, Gary E. Laderach, Daniel Max, Richard C. Pais, David P.-Y. Qing, Rachael Y. Roberts, Amanda R. Smith, Adam Starkey, Poonam Batra, Philip Caligiuri, Ali Farooqi, Gregory W. Gladish, C. Matilda Jude, Reginald F. Munden, Iva Petkovska, Leslie E. Quint, Lawrence H. Schwartz, Baskaran Sundaram, Lori E. Dodd, Charles Fenimore, David Gur, Nicholas Petrick, John Freymann, Justin Kirby, Brian Hughes, Alessi Vande Castele, Sangeeta Gupte, Maha Sallam, Michael D. Heath, Michael H. Kuhn, Ekta Dharaiya, Richard Burns, David S. Fryd, Marcos Salganicoff, Vikram Anand, Uri Shreter, Stephen Vastagh, Barbara Y. Croft, and

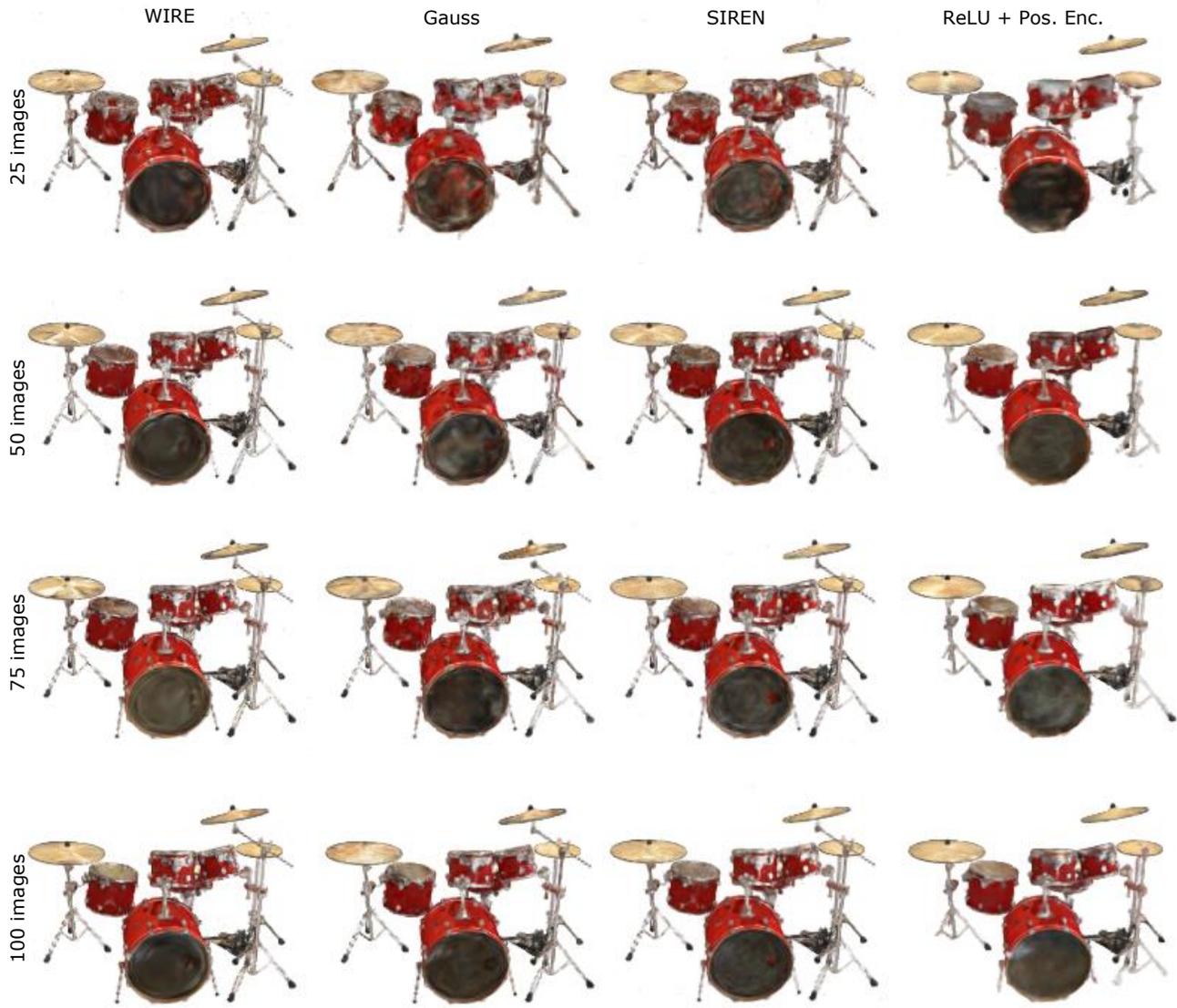


Figure 12. **Effect of number of images on NeRF accuracy.** The figure above visualizes rendered images with a neural radiance field for various nonlinearities (across columns) and varying number of images (across rows). WIRE achieves visually better reconstruction than all other methods for all numbers of images, thanks to its strong inductive biases that enable learning the high frequency features of the scene’s radiance field.

Laurence P. Clarke. The lung image database consortium (licd) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics*, 38(2):915–931, 2011. 4

[3] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *IEEE Intl. Conf. Image Processing (ICIP)*, 2007. 4

[4] Rizal Fathony, Anit Kumar Sahu, Devin Willmott, and J Zico Kolter. Multiplicative filter networks. In *Intl. Conf. Learning Representations*, 2020. 3

[5] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Adv. Neural Info. Processing Systems*, 2020. 1

[6] Jiaxiang Tang. Torch-ngp: a pytorch implementation of instant-ngp, 2022. <https://github.com/ashawkey/torch-ngp>. 5

[7] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, 2018. 2