WIRE: Wavelet Implicit Neural Representations

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan,
Ashok Veeraraghavan, Richard G. Baraniuk
Rice University
https://vishwa91.github.io/wire

Abstract

Implicit neural representations (INRs) have recently advanced numerous vision-related areas. INR performance depends strongly on the choice of activation function employed in its MLP network. A wide range of nonlinearities have been explored, but, unfortunately, current INRs designed to have high accuracy also suffer from poor robustness (to signal noise, parameter variation, etc.). Inspired by harmonic analysis, we develop a new, highly accurate and robust INR that does not exhibit this tradeoff. Our **Wavelet Implicit neural REpresentation (WIRE)** *uses as its activation function the* **complex Gabor wavelet** *that is well-known to be optimally concentrated in space–frequency and to have excellent biases for representing images. A wide range of experiments (image denoising, image inpainting, super-resolution, computed tomography reconstruction, image overfitting, and novel view synthesis with neural radiance fields) demonstrate that WIRE defines the new state of the art in INR accuracy, training time, and robustness.*

1. Introduction

Implicit neural representations (INRs), which learn a continuous function over a set of points, have emerged as a promising general-purpose signal processing framework. An INR consists of a multilayer perceptron (MLP) combining linear layers and element-wise nonlinear activation functions. Thanks to the MLP, INRs do not share the locality biases that limit the performance of convolutional neural networks (CNNs). Consequently, INRs have advanced the state of the art in numerous vision-related areas, including computer graphics [22, 27, 28], image processing [10], inverse problems [42], and signal representations [41].

Currently, INRs still face a number of obstacles that limit their use. First, for applications with high-dimensional data such as 3D volumes, training an INR to high accuracy can still take too long (tens of seconds) for real time applications. Second, INRs are not robust to signal noise or insufficient measurements. Indeed, most works on INRs in the literature assume virtually no signal noise and large amounts of data. We find in our own experiments that current INR methods are ineffective for tasks such as denoising or super-resolution. Finally, INRs still have room for improvement in representational accuracy for fine details.

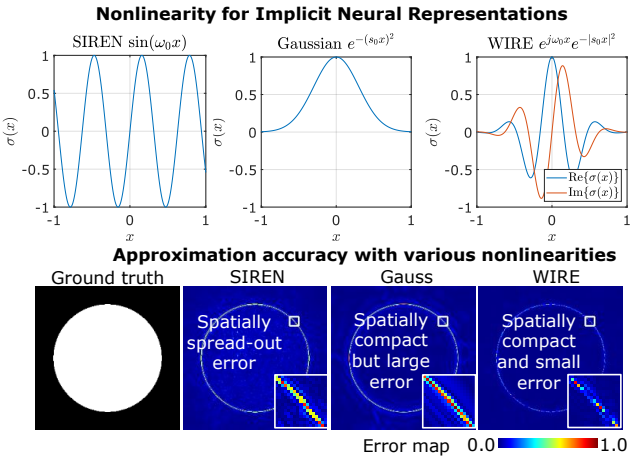


Figure 1. **Wavelet implicit neural representation (WIRE).** We propose a new nonlinearity for implicit neural representations (INRs) based on the continuous complex Gabor wavelet that has high representation capacity for visual signals. The top row shows two commonly used nonlinearities: SIREN with sinusoidal nonlinearity and Gaussian nonlinearity, and WIRE that uses a continuous complex Gabor wavelet. WIRE benefits from the frequency compactness of sine, and spatial compactness of a Gaussian nonlinearity. The bottom row shows error maps for approximating an image with strong edges. SIREN results in global ringing artifacts while Gaussian nonlinearity leads to compact but large error at edges. WIRE produces results with the smallest and most spatially compact error. This enables WIRE to learn representations accurately, while being robust to noise and undersampling of data.

In this paper, we develop a new, faster, more accurate, and robust INR that addresses these issues and takes INR performance to the next level. To achieve this, we take inspiration from harmonic analysis and reconsider the nonlinear activation function used in the MLP. Recent work has shown that an INR can be interpreted as a structured signal

representation dictionary [51], where the activation nonlinearity dictates the atoms of the dictionary. For example, the sine activation creates a pseudo-Fourier transform representation of the signal that is maximally concentrated in the frequency domain [51].

An important conclusion one can draw from the past four decades of harmonic analysis research is that Fourier methods are suboptimal for representing the kinds of signals that feature in typical vision tasks [24]. These kinds of signals, e.g., natural images from photographs, are much more concisely and robustly represented using *wavelet* atoms that are optimally concentrated in space–frequency. Sparse compositions of wavelet atoms are known to have excellent biases for representing images; cf. the seminal work in computer vision (e.g., Laplacian pyramid), computational neuroscience [30], and the JPEG2000 compression standard.

In this paper, we introduce **Wavelet Implicit neural REpresentation (WIRE)**, a new INR based on a *complex Gabor wavelet* activation function (see Figure 1). Through a wide range of experiments, we demonstrate that WIRE defines the new state of the art in INR accuracy, training time, and robustness. We showcase that WIRE's increased robustness is particularly useful for solving difficult vision inverse problems, including image denoising (robustness), image inpainting and super-resolution (superior interpolation), and 2D computed tomography (CT) reconstruction (solving higher-dimensional inverse problems). WIRE also outperforms other INRs for signal representation tasks such as overfitting images and learning point cloud occupancy volumes. Finally, we show that WIRE enables faster, more robust novel view synthesis with neural radiance fields (NeRF) [27] from critically few training views.

2. Prior Work

Regularization for inverse problems. Inverse problems involve estimating a signal from a linear or nonlinear set of measurements. Inevitably, the measurements are degraded by noise (such as camera readout or photon noise), or the problem is ill-conditioned, necessitating regularization. There are many forms of regularization, including ridge regression, Lasso [45], total variation (TV) [9], and sparsity-based [7] techniques that seek to penalize the ℓ_1 norm the signal or some transform thereof. In the past decade, data-driven regularization, including overcomplete dictionary-based [4] and generative network-based [29, 35, 36] techniques, have been developed. The classical model-based approaches are inadequate for severely ill-conditioned problems, while the data-driven ones critically depend on data.

Convolutional neural networks (CNNs). CNNs, the most popular neural network architectures in computer vision for the past decade, have been shown to exhibit strong implicit biases that favor image-like signals. This has been demonstrated with works like deep image prior (DIP) [47] and its variations [13, 19] that produce remarkable results on image-related linear inverse problems without any prior training data. However, such CNN-based priors are tied to a discrete grid-like signal representation which is not applicable to problems such as novel view synthesis, or for solving ordinary and partial differential equations, and not scalable for very high dimensional signals such as 3D tomographic volumes, gigapixel images, or large point clouds.

Deep image prior. Neural networks, and particularly CNNs, exhibit implicit biases due to their specific architectures (such as a UNet [37]), implying that even untrained neural networks can be used for regularization. This was leveraged to build a deep image prior (DIP) [47] that produces outputs that tend to look like images. While DIP offers superior regularization to analytical approaches, it often exhibits good performance only when over-parameterized and are tied to a grid-like discretized representation of the signal, implying DIPs do not scale to high dimensional signals such as point clouds with a large number of points. The issue of computational cost has been addressed to a certain extent by the deep decoder [19] and the DeepTensor [38], but they still need the signal to be defined as a regular data grid such as a 2D matrix or 3D tensor.

Implicit representations. INRs are continuous learned function approximators based on multilayer perceptrons (MLPs). The continuous nature of INRs is particularly appealing when dealing with irregularly sampled signals such as a point clouds. Since its first widespread usage in novel view synthesis in graphics [27], INRs have pervaded nearly all fields of vision and signal processing including rendering [22], computational imaging [5, 11], medical imaging [49], and virtual reality [14].

The popular choice of the ReLU nonlinearity in standard neural networks has been empirically shown to result in poor approximation accuracy in INRs. This has been remedied by several modifications to the MLP including the so-called positional encoding [28, 43], as well as various choices of nonlinearity such as the sinusoidal function [41] and the Gaussian function [33]. A closely related work is the Gabor wavelet-based multiplicative filter networks (MFN), where the output after each layer is multiplied by a Gabor filter. The output then results in a combination of exponentially many Gabor wavelets, thereby resulting in large capacity. Numerous architectural changes have also been proposed that leverage multiscale properties of visual signals to accelerate the INR training procedure including adaptive block decomposition [25], kilo-NeRF [34], and predicting the Laplacian pyramid of the signal [39].

INRs can now train on signals nearly instantly [28] thanks to these numerous advances. However, the high ca-

pacity of such INRs precludes robustness — implying that the signal representation is brittle, resulting in overfitting to both noise and signal equally. In this paper, we propose the complex Gabor wavelet as a nonlinearity, which is uniquely well-suited to induce robustness in INRs.

Wavelet transform and the Gabor wavelet. The Fourier transform decomposes the signal as a sum of sinusoids with infinite space support, implying that there is no notion of spatial compactness. The wavelet transform remedies this by decomposing the signal as a linear combination of translated and scaled versions of a short oscillating pulse called a wavelet. Wavelets typically result in faster approximation rates for signals and images than Fourier transform [15], and hence they are often used for image compression [16, 40] and as a robust prior for inverse problems of images [18] and videos [48]. In this paper, we show that wavelets are a universally superior choice for the nonlinearity in INRs due to their compact support in space and frequency and therefore faster approximation rates.

3. Wavelet Implicit Representations

3.1. INR details

Consider an INR function $F_\theta : \mathbb{R}^{D_i} \mapsto \mathbb{R}^{D_o}$ mapping D_i input dimensions to D_o output dimensions, where θ represents the MLP's tunable parameters. The goal is to construct F_θ such that it approximates a function $g(\mathbf{x})$ of interest, i.e., $g(\mathbf{x}) \approx F_\theta(\mathbf{x})$. For example, $g(\mathbf{x})$ may simply be a ground truth image, represented as a function mapping coordinates to pixel values. Modeling $F_\theta(\cdot)$ as an M-layer MLP, the output at each layer is given by

$$\mathbf{y}_m = \sigma(W_m \mathbf{y}_{m-1} + \mathbf{b}_m), \tag{1}$$

where σ is the nonlinearity (or nonlinear activation function); $W_m, \mathbf{b}_m$ are weights and biases for the m^{th} layer; $\mathbf{y}_0 = \mathbf{x} \in \mathbb{R}^{D_i}$ is the input coordinate and $\mathbf{y}_{M+1} = W_{M+1}\mathbf{y}_M + \mathbf{b}_{M+1}$ is the final output.

The nonlinear activation σ plays a key role in the representation capacity of the INR (see Fig. 1). Two leading choices include the periodic $\sigma(x) = \sin(\omega_0 x)$ used in SIREN [41], and the Gaussian nonlinearity $\sigma(x) = e^{-(s_0 x)^2}$ used by Ramasinghe et. al. [33]—both result in significantly higher representation accuracy than ReLU. However, their high representation capacity is also a drawback, since they can represent noise with nearly equal accuracy as an image. Our goal is to propose a nonlinearity σ that is well-suited for visual signals such as images, videos, and 3D volumes but poorly fits noise-like signals.

3.2. WIRE

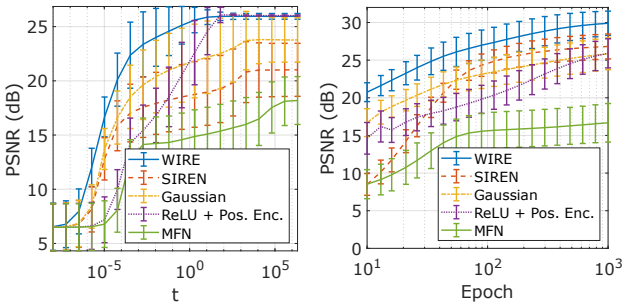
Armed with the insight that a Gabor wavelet achieves optimal time-frequency compactness, we propose the wavelet implicit representations (WIRE) that use the **continuous complex Gabor wavelet ψ as a nonlinearity**:

$$\sigma(x) = \psi(x; \omega_0, s_0) = e^{j\omega_0 x} e^{-|s_0 x|^2}, \tag{2}$$

where ω_0 controls the frequency of the wavelet and s_0 controls the spread (or width). The first layer activations have the form

$$\mathbf{y}_1 = \psi(W_1 \mathbf{x} + \mathbf{b}_1; \omega_0, s_0), \tag{3}$$

which are copies of the mother Gabor wavelet ψ at scales and shifts determined by W_1 and $\mathbf{b}_1$. Hence the building blocks of WIRE are drawn from a dictionary of wavelet functions. We let the weights of the INR as well as the outputs be complex-valued to preserve phase relationships throughout, and we represent real signals by simply taking the real part of the output and discarding the imaginary part. Just as wavelets combine space and frequency compactness, WIRE enjoys the advantages of periodic nonlinearities such as SIREN due to the complex exponential term and the spatial compactness from the Gaussian window term; recall Fig. 1. Additionally, unlike SIREN, WIRE does not require a carefully chosen set of initial weights (see Fig. 3) due to the Gaussian window, which creates a spatially compact output at each layer and produces high quality results with the default neural network initialization of uniformly random weights independent of the parameters ω_0, s_0.



(a) Implicit bias with NTK gradient flow

(b) Implicit bias in standard INRs.

Figure 2. **Implicit bias in denoising** (a) The empirical NTK of finite-width INRs provides an insight into the implicit bias of INRs. Early trajectories of NTK gradient flow show WIRE converging to the image faster than the noise, outperforming all other nonlinearities. Bars indicate one standard deviation over the dataset. (b) Early iterations of standard training are reflected well by the relative performances of NTK gradient flow from part (a). Furthermore, WIRE maintains its advantage against other nonlinearities throughout the remainder of training.

3.3. Implicit bias of WIRE

Neural tangent kernel perspective. As stated, we seek an INR that fits visual signals well but fits noise poorly in

comparison. Inspired by [43, 51], who proposed to compare eigenfunctions of the empirical neural tangent kernel (NTK) [20] of INRs to understand their approximation properties, we compare the fitting of noisy natural images using NTK gradient flow. The NTK gradient flow of INRs accurately captures the behavior of early training of neural networks, and so in tasks such as denoising where we regularize via early stopping, the early training behavior determines the implicit bias. In the lazy training regime of wide neural networks [23], the fit image at time $t \geq 0$ has value

$$F_{\theta_t}(\mathbf{x}) = [(I - e^{-tK})g](\mathbf{x}), \tag{4}$$

where I is the identity operator, K is the NTK operator on the image's spatial domain, and g is the image being fit.

In Fig. 2a, we apply NTK gradient flow using the empirical finite-width NTK to a denoising task, fitting the original image with $\mathcal{N}(0, 0.05^2)$ i.i.d. pixel-wise additive noise. We evaluate on $64 \times 64 \times 3$ images from Tiny ImageNet [2]. Comparing WIRE to other INRs, we see that, WIRE prefers to learn the signal in the image early in training rather than the noise, converging orders of magnitude faster to essentially any given peak signal-to-noise-ratio (PSNR).

Empirical evaluation. We perform an anlogous denoising task for real INRs on the 24 $768 \times 512 \times 3$ images from the Kodak Lossless True Color Image Suite [1], again with $\mathcal{N}(0, 0.05^2)$ additive noise, in Fig. 2b. We apply the same INRs as in the NTK example, but train with ordinary neural network gradient optimization instead of NTK gradient flow. Again, WIRE drastically outperforms other INRs, converging an order of magnitude faster to the same PSNR.

3.4. Choosing the parameters ω_0, s_0

WIRE's performance is primarily decided by the constants ω_0, s_0 that control frequency of the sinusoid and width of the Gaussian, respectively. WIRE outperforms both the SIREN and Gaussian nonlinearities across a broad range of parameters. Figure 3 shows the approximation accuracy achieved by WIRE for various parameters. We set the number of hidden layers to three, and number of hidden features to 256. When $\omega_0 = 0$, we used a Gaussian nonlinearity. When $s_0 = 0$, we used a sinusoidal nonlinearity. When both parameters were zero, we used a ReLU nonlinearity. For the denoising task, we added photon noise equivalent to a maximum of 50 photons per pixel. We observe from Fig. 3 that WIRE outperforms SIREN, Gauss, and ReLU. Moreover, the performance is superior for a large swath of values of ω_0, s_0 for both image representation and denoising. The reduced sensitivity to the exact values of ω_0, s_0 implies that WIRE can be used without precise information about image or noise statistics. Additional experiments on sensitivity to parameters is in the supplementary.

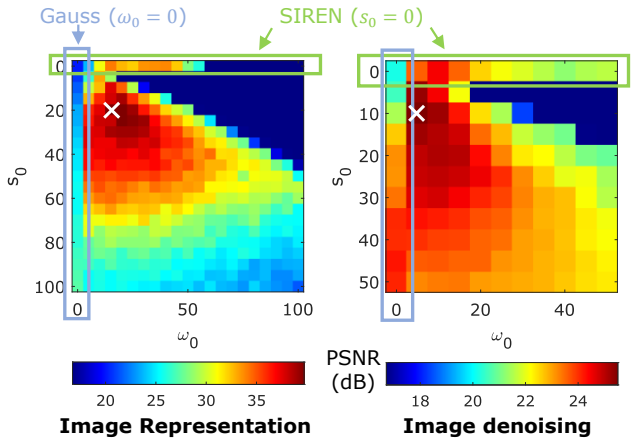
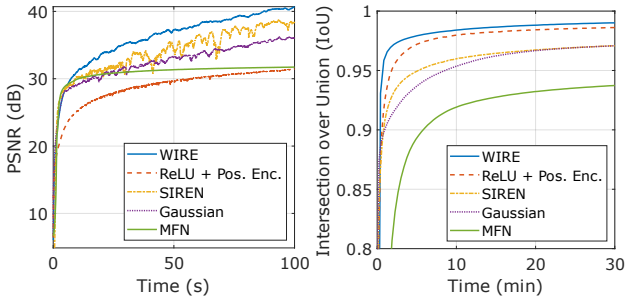


Figure 3. **WIRE is robust to the choice of parameters.** The plot above shows accuracy for image representation and denoising with various settings of ω_0 and s_0. The boxes show special cases with $\omega = 0$ corresponding to Gaussian nonlinearity, and $s_0 = 0$ corresponding to SIREN. WIRE achieves higher accuracy than both SIREN and Gauss on image representation as well as image denoising tasks (marked by white cross). Further, WIRE achieves super performance for a large choice of parameters ω_0, s_0 implying that WIRE is not overly sensitive to the hyperparameters.

Alternate forms of WIRE. For problems where complex weights are infeasible, WIRE can be instantiated as the imaginary (or real) part of the complex Gabor wavelet, $\psi(x; \omega_0, s_0) = \sin(\omega_0 x) e^{-(s_0 x)^2}$. Note that setting $s_0 = 0$ results in the sine nonlinearity used in SIREN [41] and $\omega_0 = 0$ results in Gaussian nonlinearity [33] implying WIRE inherits the favorable properties of previously proposed nonlinearities. Another embodiment of WIRE is a Constant-Q Gabor wavelet where $\omega_0 s_0 = Q$, which results in constant fractional bandwidth $(\omega/\delta\omega)$. Constant-Q Gabor wavelets are often used in music analysis [8, 46], wavelet transforms [24], and the Laplace transform. Having only a single parameter makes hyperparameter tuning simpler with a fixed Q. Performance with real Gabor, and other continuous wavelets is in the supplementary.

4. Experiments

WIRE learns representations for all signal classes faster than state-of-the-art techniques. In addition, WIRE is well-suited to solve a large class of inverse problems with limited measurements, or when the measurements are corrupted by noise. For all the experiments, we implemented the optimization in PyTorch [31] and used the Adam optimizer [21]. Code was executed on a system unit equipped with 64GB RAM, and an Nvidia RTX 2080 Ti graphical processing unit (GPU) with 8GB memory. We used an ℓ_2 loss function between the measurements and the outputs of INR with no further regularization. We scheduled the learning rate to de-



(a) Image representation (b) Volume representation

Figure 4. **WIRE learns faster.** The two plots above show representation accuracy for an image (top row in Fig. 5) and an occupancy volume (bottom row in Fig. 5) over time. Owing to the high approximation capacity of Gabor wavelets for visual signals, WIRE achieves high accuracy at a faster rate, making it an appropriate choice for representing visual signals.

cayed the initial learning rate by 0.1 at the end of training.

4.1. Signal representation

We evaluate WIRE on representing images and occupancy volumes [26]. In both cases, we used an MLP with three hidden layers with a width of 300 features for all nonlinearities. For WIRE, we reduced the number of parameters by half by reducing the hidden features by $\sqrt{2}$ to account for the doubling due to real and imaginary parts. The parameters for each nonlinearity and the learning rate were chosen to obtain fastest approximation rate. Specifically, we chose $\omega_0 = 20, s_0 = 10$ for WIRE, $\omega_0 = 40$ for SIREN, and $s_0 = 30$ for Gaussian. We also compare against multiplicative frequency networks (MFN) [17]. For the occupancy volume, we sampled over a $512 \times 512 \times 512$ grid with each voxel within the volume assigned a 1, and voxels outside the volume assigned a 0. We evaluated the PSNR and structural similarity (SSIM) [50] for images and intersection over union (IOU) for the occupancy volumes.

Figure 4 shows the approximation accuracy as a function of time for an image (Kodak dataset) and an occupancy volume (Thai statue). WIRE achieves the highest accuracy, while converging faster than other approaches. Figure 5 visualizes the final representation of the example image after 1.6 minutes, and the 3D mesh of the Thai Statue constructed with marching cubes after 30 minutes. WIRE achieves the highest accuracy both for images (43.2dB) and for the occupancy volume (0.99), confirming our hypothesis that WIRE has higher approximation accuracy.

4.2. Solving inverse problems of 2D images

WIRE's inductive bias favors images, and hence can be used for solving linear inverse problems. To demonstrate the advantages of WIRE as a strong prior for images, we showcase its performance on image denoising, single image super resolution, and multi-image super resolution.

Image denoising. To evaluate the robustness of INRs for representing noisy signals, we learned a representation on a high resolution color image from the DIV2K dataset [3]. We simulated photon noise with an independently distributed Poisson random variable at each pixel with a maximum mean photon count of 30, and a readout count of 2, resulting in an input PSNR of 17.6 dB. We then learned a representation on this noisy image with various nonlinearities. In all cases, we chose an MLP with two hidden layers and 256 features per layer. We also compared the denoising result with state-of-the-art self-supervised denoising technique, "self2self [32]. Figure 6 visualizes the final result for each nonlinearity along with metrics for each result. WIRE produces the sharpest image with least amount of residual noise. Qualitatively, WIRE's result is similar to that by self2self, implying that WIRE enjoys inductive biases that make it a good choice for inverse problems.

Image super resolution. INRs function as interpolators, and hence super resolution benefits from INRs with good implicit biases. We test this by implementing $4\times$ super resolution on a DIV2K image. The forward operator can be cast as $\mathbf{y} = A_4\mathbf{x}$ where A_4 implements a $4\times$ downsampling operator (without aliasing). We then solved for the sharp image by modeling $\mathbf{x}$ as output of an INR. Figure 7 visualizes the result on super resolution of image of a butterfly with various approaches. WIRE produces the sharpest result with crisp details on the butterfly's antenna and on the wings. WIRE results are similar to the deep image prior (DIP) [47], establishing the generality of WIRE.

INRs are particularly advantageous when data interpolation needs to be performed on an irregular grid. An example of such settings is multi-image super resolution where the images are shifted and rotated with respect to each other. Figure 8 shows an example of $4\times$ super resolution with four images (and hence 25% compression) from the Kodak dataset [1] simulated with a small sub-pixel motion between them. The forward operator is then $\mathbf{y}^k = A_4^k\mathbf{x}$ where A_4^k encodes the downsampling, and translation and rotation for the k^{th} image. The visualizations in the figure demonstrate that WIRE achieves the highest accuracy and is qualitatively better at reconstructing high frequency components. In contrast, the Gaussian nonlinearity leads to a blurry reconstruction, while SIREN results in ringing artifacts.

Computed tomography (CT) reconstruction. Strong signal priors are critical for solving underconstrained problems, and CT reconstruction is one such example. We emulated 100 CT measurements of a 256×256 x-ray colorectal image [12]. Figure 9 shows the final reconstruction with

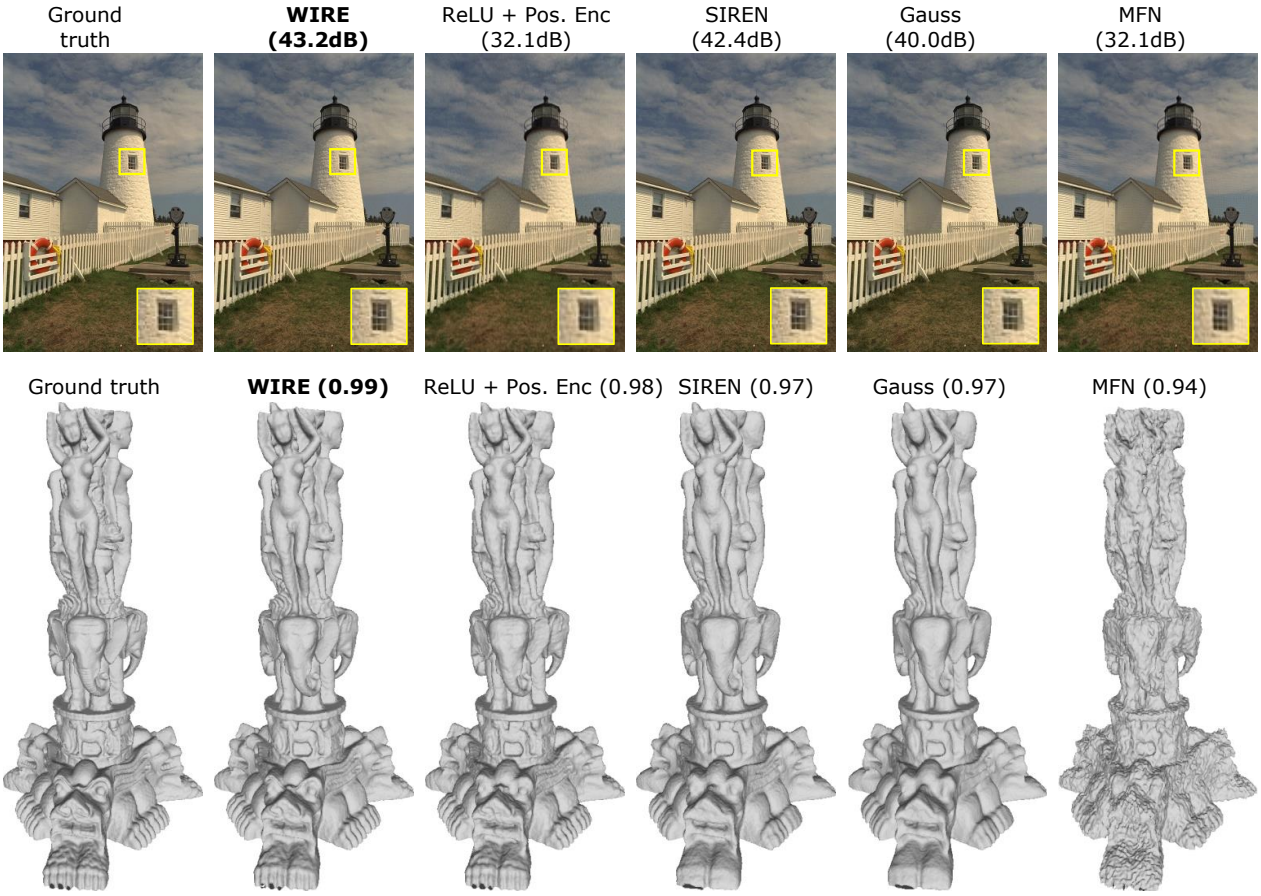


Figure 5. **WIRE has high representation capacity.** The results above show image representation in the first row and meshes generated with occupancy volumes in the second row with various nonlinearities. WIRE achieves highest representation accuracy for both data, underlining its advantages as a signal model.

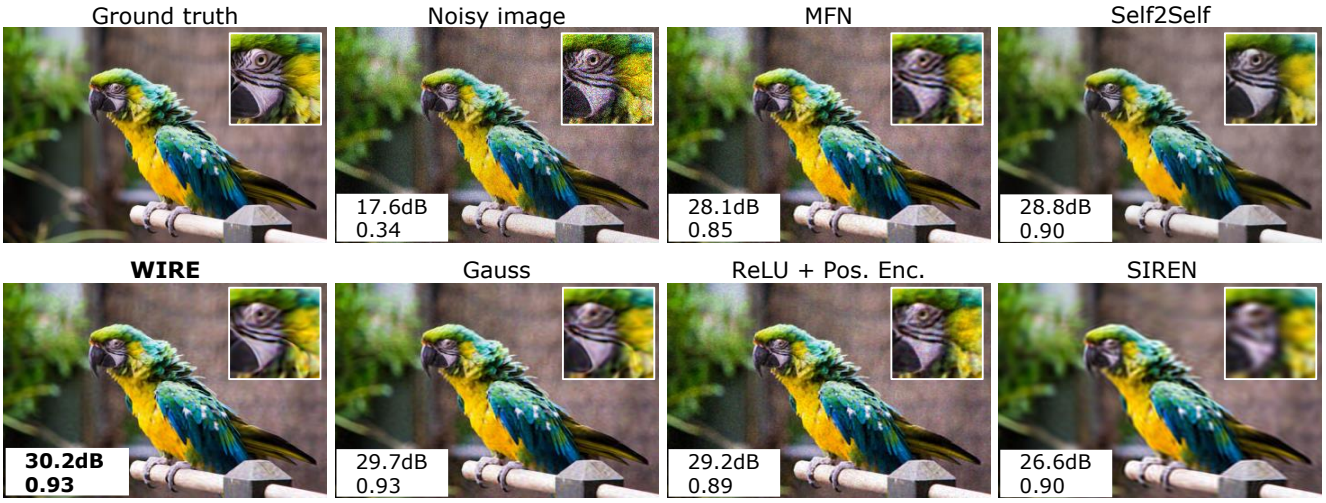


Figure 6. **WIRE is robust to noise.** We show an image representation with added shot noise, resulting in an input PSNR of 17.6dB. Among the various approaches, WIRE results in the highest PSNR and SSIM of any representation. Further, WIRE results in similar or better denoising results compared to state-of-the-art self-supervised denoising approaches such as self2self [32].

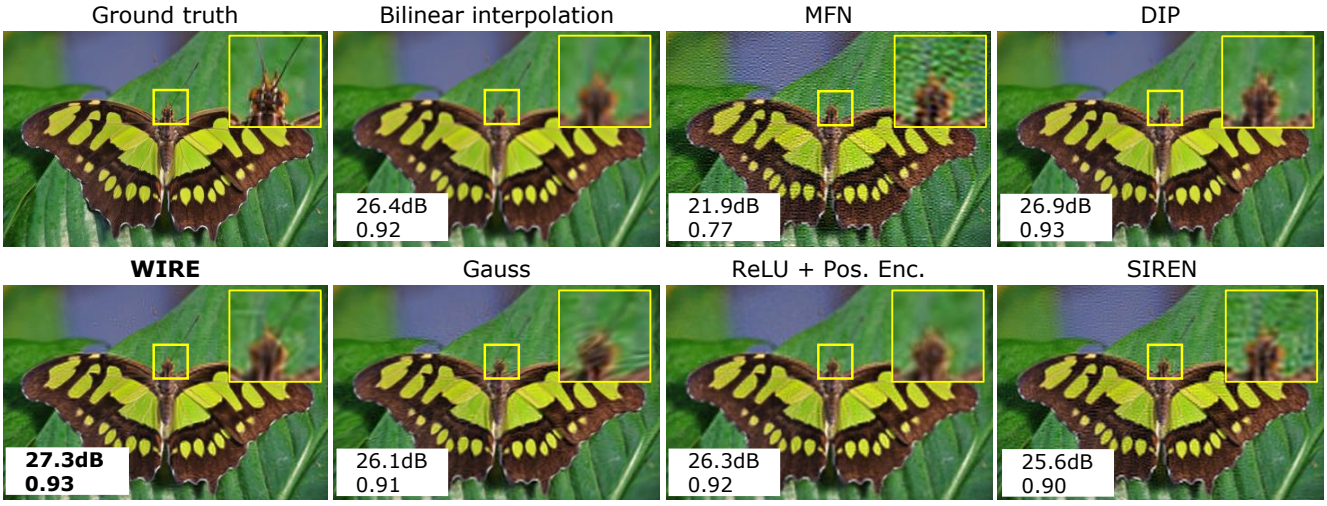


Figure 7. **WIRE for single image super resolution.** The figure above shows results for a $4\times$ single image super resolution with various approaches. Thanks to its strong implicit bias, WIRE results in the sharpest reconstruction with quantitatively higher reconstruction metrics.

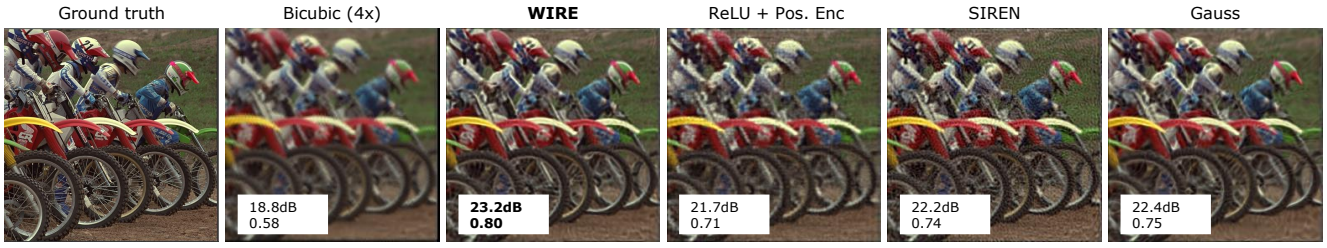


Figure 8. **Multi-image super resolution.** INRS are particularly appealing for handling data on an irregular grid, such as images captured with multiple sub-pixel shifts. The figure above shows $4\times$ super resolution with 4 images captured with varying sub-pixel shifts and rotations. We then solved a joint inverse problem where the high resolution image is modeled as the output of an INR. WIRE produces the best reconstruction both quantitatively and qualitatively, implying that WIRE has favorable interpolation properties for visual signals.

various approaches. WIRE results in the sharpest reconstruction with clearly pronounced features. SIREN has striation artifacts that are expected from an unregularized reconstruction while Gaussian nonlinearity results in overly smooth results. WIRE can hence be used as a robust prior for inverse problems with noisy and undersampled measurements.

4.3. Learning neural radiance fields

INRs have been leveraged successfully for novel-view synthesis with neural radiance fields (NeRF) [27]. Given images from a sparse set of view points, the goal is to render an image from a novel view. NeRF achieves this by training a common INR three positional and two directional inputs, and produces transmission and color at that location. Images are then produced by integrating along lines that pass through each view's lens (pinhole). The simplest NeRF architecture consists of positional encoding, and two MLPs equipped with ReLU for transmission and color values. We show that WIRE without any positional encoding produces higher quality results within fewer epochs.

We trained NeRFs for reconstruction on the synthetic drum dataset [27]. Each image was downsampled to a resolution of 200×200. We then trained the radiance field with only 25 images instead of the default 100 images. Additional experiments with varying number of training images is in the supplementary. We used the "torch-NGP" codebase [44] for training the NeRF model. For all experiments, we chose a 4-layered MLP with a width of 128 features for each layer. Parameters were chosen to achieve fastest rate of increase of approximation accuracy on the validation dataset. Figure 10 shows results with various nonlinearities. WIRE produces highest accuracy (+0.2dB) with fastest rate of increase. WIRE learns features absent in outputs of other nonlinearities such as the rod connecting the ride cymbal to its stand and the anisotropic reflections on the cymbal.

5. Conclusions

We have proposed and validated the advantages of WIRE, an INRs equipped with a complex Gabor wavelet

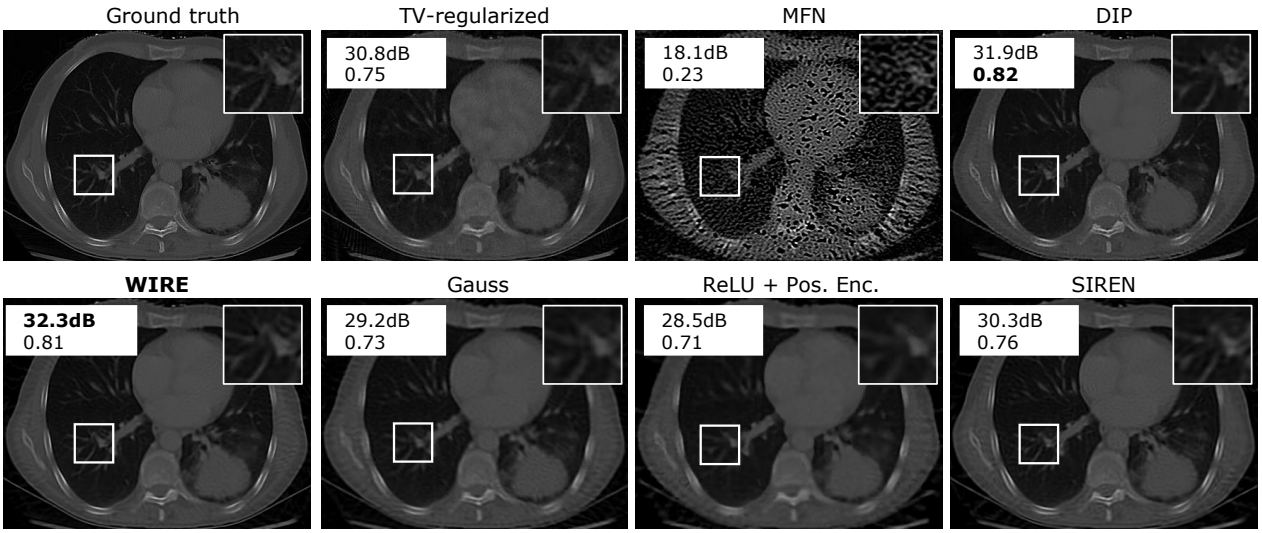


Figure 9. **Computed tomography reconstruction.** Inverse problems with noisy undersampled data require a strong signal prior for robust reconstruction. Here, we show CT-based reconstruction with 100 angles for a 256×256 image ($2.5\times$ compression) with various approaches. WIRE results in sharp reconstruction, exposing features that are blurry, or with ringing artifacts in reconstructions with other approaches. WIRE is hence a strong signal prior for images, and can solve a large class of inverse problems.

Figure 10. **Novel-view synthesis with neural radiance fields.** INRs have shown most promise in novel-view synthesis where the transmittance and color at each 3D voxel is modeled as output of INRs. Here, we show that WIRE is well-suited for novel-view synthesis with no additional positional encoding. WIRE not only achieves higher accuracy (+0.2dB) with fewer epochs, but captures details that are missed out by other nonlinearities, such as the rod connecting the ride cymbal to its stand and the anisotropic reflections on the cymbals.

nonlinear activation function. Previously proposed nonlinearities have complementary strengths: SIREN has high representation capacity and trains fast but underperforms on inverse problems. Positional encoding has lower capacity but excels at novel-view synthesis. Gaussian is favorable for denoising tasks. In contrast, on the practical side, WIRE inherits the best properties of all of the above nonlinearities, namely (i) higher representation capacity, (ii) higher accuracy at a faster rate, and (iii) a strong and realistic inductive bias for solving challenging inverse problems. On the theoretical side, WIRE provides an intriguing avenue to introduce ideas from computational harmonic analysis (i.e., multiscale wavelet analysis and optimized analyses [6]) into INRs in particular and potentially deep learning in general.

6. Acknowledgements

This work was supported by NSF grants CCF-1911094, IIS-1838177, IIS-1652633, IIS-2107313, and IIS-1730574; ONR grants N00014-18-12571, N00014-20-1-2534, N00014-18-1-2047, and MURI N00014-20-1-2787; AFOSR grant FA9550-22-1-0060; and a Vannevar Bush Faculty Fellowship.

References

[1] Kodak lossless true color image suite. http://r0k.us/graphics/kodak/, 1999. Accessed: 2022-11-09. 4, 5

[2] Tiny imagenet visual recognition challenge. https://www.kaggle.com/c/tiny-imagenet, 2015. Accessed: 2022-11-03. 4

[3] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, July 2017. 5

[4] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006. 2

[5] Benjamin Attal, Eliot Laidlaw, Aaron Gokaslan, Changil Kim, Christian Richardt, James Tompkin, and Matthew O'Toole. Törf: Time-of-flight radiance fields for dynamic scene view synthesis. *Adv. Neural Info. Processing Systems*, 34:26289–26301, 2021. 2

[6] Richard Baraniuk. Optimal tree approximation with wavelets. In *Wavelet Applications in Signal and Image Processing VII*, volume SPIE 3813, pages 196–207, 1999. 8

[7] Richard G Baraniuk, Volkan Cevher, Marco F Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Trans. Info. Theory*, 56(4):1982–2001, 2010. 2

[8] Judith C Brown. Calculation of a constant Q spectral transform. *The J. Acoustical Society of America*, 89(1):425–434, 1991. 4

[9] Antonin Chambolle. An algorithm for total variation minimization and applications. *J. Mathematical Imaging and Vision*, 20(1):89–97, 2004. 2

[10] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, 2021. 1

[11] Chiun-Hong Chien and Jake K Aggarwal. Volume/surface octrees for the representation of three-dimensional objects. *Computer Vision, Graphics, and Image Processing*, 36(1):100–113, 1986. 2

[12] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J. Digital Imaging*, 26(6):1045–1057, 2013. 5

[13] Mohammad Zalbagi Darestani and Reinhard Heckel. Accelerated MRI with un-trained neural networks. *IEEE Trans. Computational Imaging*, 7:724–733, 2021. 2

[14] Nianchen Deng, Zhenyi He, Jiannan Ye, Budmonde Duinkharjav, Praneeth Chakravarthula, Xubo Yang, and Qi Sun. Fov-nerf: Foveated neural radiance fields for virtual reality. *IEEE Trans. Visualization and Computer Graphics*, 28(11):3854–3864, 2022. 2

[15] Ronald A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998. 3

[16] Ronald A DeVore, Björn Jawerth, and Bradley J Lucier. Image compression through wavelet transform coding. *IEEE Trans. Info. Theory*, 38(2):719–746, 1992. 3

[17] Rizal Fathony, Anit Kumar Sahu, Devin Willmott, and J Zico Kolter. Multiplicative filter networks. In *Intl. Conf. Learning Representations*, 2020. 5

[18] Lihan He and Lawrence Carin. Exploiting structure in wavelet-based Bayesian compressive sensing. *IEEE Trans. Signal Processing*, 57(9):3488–3497, 2009. 3

[19] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. In *Intl. Conf. Learning Representations*, 2018. 2

[20] Arthur Jacot, Franck Gabriel, and Clement Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Adv. Neural Info. Processing Systems*, 2018. 4

[21] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Intl. Conf. Learning Representations*, 2015. 4

[22] Alexandr Kuznetsov, Krishna Mullia, Zexiang Xu, Miloš Hašan, and Ravi Ramamoorthi. NeuMIP: Multi-resolution neural materials. *ACM Trans. Graphics*, 40(4):1–13, 2021. 1, 2

[23] Jaehoon Lee, Lechao Xiao, Samuel Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In *Adv. Neural Info. Processing Systems*, 2019. 4

[24] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999. 2, 4

[25] Julien NP Martel, David B Lindell, Connor Z Lin, Eric R Chan, Marco Monteiro, and Gordon Wetzstein. Acorn: Adaptive coordinate networks for neural scene representation. *arXiv preprint arXiv:2105.02788*, 2021. 2

[26] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3D reconstruction in function space. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, 2019. 5

[27] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *IEEE European Conf. Computer Vision (ECCV)*, 2020. 1, 2, 7

[28] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, 2022. 1, 2

[29] Anh Nguyen, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, and Jason Yosinski. Plug & play generative networks: Conditional iterative generation of images in latent space. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, 2017. 2

[30] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996. 2

[31] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In

Adv. Neural Info. Processing Systems. 2019. 4

[32] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, 2020. 5, 6

[33] Sameera Ramasinghe and Simon Lucey. Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In *IEEE European Conf. Computer Vision (ECCV)*, 2021. 2, 3, 4

[34] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. *arXiv preprint arXiv:2103.13744*, 2021. 2

[35] JH Rick Chang, Chun-Liang Li, Barnabas Poczos, BVK Vijaya Kumar, and Aswin C Sankaranarayanan. One network to solve them all–solving linear inverse problems using deep projection models. In *IEEE Intl. Conf. Computer Vision (ICCV)*, 2017. 2

[36] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM J. Imaging Sciences*, 10(4):1804–1844, 2017. 2

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Intl. Conf. Medical Image Computing and Computer-Assisted Intervention*, 2015. 2

[38] Vishwanath Saragadam, Randall Balestriero, Ashok Veeraraghavan, and Richard G Baraniuk. Deeptensor: Low-rank tensor decomposition with deep network priors. *arXiv preprint arXiv:2204.03145*, 2022. 2

[39] Vishwanath Saragadam, Jasper Tan, Guha Balakrishnan, Richard Baraniuk, and Ashok Veeraraghavan. Miner: Multiscale implicit neural representations. In *European Conf. Computer Vision*, 2022. 2

[40] Jerome M Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Signal Processing*, 41(12):3445–3462, 1993. 3

[41] 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, 2, 3, 4

[42] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for imaging inverse problems. *arXiv preprint arXiv:2102.05181*, 2021. 1

[43] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Adv. Neural Info. Processing Systems*, 2020. 2, 4

[44] Jiaxiang Tang. Torch-ngp: a pytorch implementation of instant-ngp, 2022. https://github.com/ashawkey/torch-ngp. 7

[45] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. 2

[46] Massimiliano Todisco, Héctor Delgado, and Nicholas Evans. Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification. *Computer Speech & Language*, 45:516–535, 2017. 4

[47] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, 2018. 2, 5

[48] Michael B. Wakin, Jason N. Laska, Marco F. Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin F. Kelly, and Richard G. Baraniuk. Compressive imaging for video representation and coding. In *PCS Proc.*, 2006. 3

[49] Yuehao Wang, Yonghao Long, Siu Hin Fan, and Qi Dou. Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery. In *Intl. Conf. Medical Image Computing and Computer-Assisted Intervention*, 2022. 2

[50] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, 2004. 5

[51] Gizem Yüce, Guillermo Ortiz-Jiménez, Beril Besbinar, and Pascal Frossard. A structured dictionary perspective on implicit neural representations. In *IEEE Comp. Vision and Pattern Recognition (CVPR)*, 2022. 2, 4