CoRe-INR: Coordinate Rescaling for Spectral Expansion and Convergence Improvement for Implicit Neural Representations

Anonymous authors
Paper under double-blind review

Abstract

Implicit Neural Representations (INRs) create continuous representations for a broad range of signals by mapping the signal coordinates and corresponding values. However, INRs have limitations in learning high-frequency components due to the spectral bias. In this research, we explore the advantage of coordinate rescaling to mitigate spectral bias. We theoretically explain that scaling would improve the spectral representation characteristics using frequency atoms with a trade-off of introducing a spectral jitter. This spectral-aligned scaling improves convergence, explained through the Neural Tangent Kernel and the novel perspective of the spectral representation of tangent features. Specifically, we observe that some signals demonstrate a delayed learning behavior while scaling reduces this barrier to reach high-frequency content, explained through the tangent features. Further, through experimental results, we argue that there is a potential to outperform a deeper network with a shallow network, with the support of scaling. Moving further, we introduce CoRe-INR that demonstrates axis-based scaling could improve representation and inverse task modeling of high-dimensional signals.CoRe-INR outperforms the state-of-the-art algorithms, notably obtaining a +1.7 dB improvement in image representation. Our groundwork provides a new opening for future research in this domain.

1 Introduction

Implicit neural representations (INRs) aims to model a signal as a continuous coordinate-conditioned function and have demonstrated promising results in signal representationSitzmann et al. (2020); Mescheder et al. (2019) and many other inverse tasks Sitzmann et al. (2019); Mildenhall et al. (2021); Vyas et al. (2025). However, these INR networks typically use feed forward multilayer perceptron networks and they exhibit spectral bias. This is the effect where the network learns low-frequency structures more easily and promptly than high-frequency details Rahaman et al. (2019).

Previous works uses fourier features to mitigate this limitation by introducing a prescribed sinusoidal encoding Tancik et al. (2020), while SIREN employs periodic activations to represent high-frequency signals and their derivatives Sitzmann et al. (2020). Subsequent methods modify the encoding, activation, or network parameterisation to improve spectral coverage Saragadam et al. (2023); Liu et al. (2024); Kazerouni et al. (2024); Shi et al. (2024); Jayasundara et al. (2025); Thennakoon et al. (2026). Nevertheless, the effect of directly scaling the coordinate domain remains poorly explored.

The authors of SIREN expands the input time axis for high-rate audio and controls the first-layer frequency through ω_0 Sitzmann et al. (2020). Although coordinate scaling can be absorbed into the first-layer transformation, it is typically treated as an application-specific normalisation choice. We instead study it as a mechanism that changes both the spectral content represented by a sinusoidal network and the dynamics by which that content is learned. Experiments on audio signals show that larger scales broaden the effective hidden-layer spectra and shorten the initial PSNR plateau.

Under the transformation $\boldsymbol{r} \mapsto \mathbf{A}\boldsymbol{r}$, a frequency atom $\boldsymbol{\omega}$ becomes $\mathbf{A}^{\mathsf{T}}\boldsymbol{\omega}$. Building on the structured dictionary interpretation of sinusoidal INRs Yüce et al. (2022), this transformation applies to both first-layer frequencies and the sum-frequency atoms generated through deeper sinusoidal com-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

positions. There we can say that scaling input coordinates dilates the harmonic dictionary. This scaling also modifies the tangent features and empirical NTK, enabling faster convergence when target-relevant modes obtain larger eigenvalues and stronger alignment.

Our observations reveal a scale depth trade-off: coordinate scaling can shift the available frequency atoms towards the target spectrum without adding layers, allowing a suitably scaled shallow SIREN to outperform a deeper, larger unscaled model. Motivated by this finding, we introduce **CoRe-INR**, a multi-branch INR using the scaling configurations $(1, 1)$, $(s_x, 1)$, $(1, s_y)$, and (s_x, s_y). Each branch predicts a signal value and a confidence logit, and the final output is formed as a spatially varying softmax-weighted combination of all branch predictions.

The main contributions of this work are a theoretical analysis of coordinate scaling and an empirical evaluation of the scale depth trade-off.

- We explain theoretically and experimentally of the advantage of input coordinate scaling using frequency atom dilation, showing that there exists a trade-off for scaling, and a shallow INR can achieve a similar performance in spectral characteristics to an unscaled deeper network.
- We provide a learning dynamic based explanation of coordinate scaling in INRs through a novel approach of the spectral representation of the Tangent Features that explains the delayed learning behavior in some signal representation tasks.
- We introduce CoRe-INR, an axis-rescaled multi-branch architecture with confidence-based fusion, and demonstrate its effectiveness for audio, image, and three-dimensional object representation, as well as inverse imaging problems.

2 Related Work

Frequency and scale control in INRs. Frequency control has consequently been introduced at several locations in the INR corpus. Positional and Fourier encodings scale the frequencies supplied to the network Tancik et al. (2020), while SIREN controls its first-layer spectrum through ω_0 and expands the coordinate domain for high-rate audio Sitzmann et al. (2020). Other methods redesign the activation function, including non-periodic coordinate-MLP activations Ramasinghe & Lucey (2022), Gabor-wavelet activations in WIRE Saragadam et al. (2023), variable-periodic activations in FINER Liu et al. (2024), conditioned sinusoidal parameters in INCODE Kazerouni et al. (2024), and complex sinusoidal modulation in COSMO-INR Thennakoon et al. (2026). Beyond input encodings and activation design, recent methods control INR spectra through Fourier-reparameterized weights, signal-matched activation atoms, adaptive sinusoidal parameters, and scaled initialization Shi et al. (2024); Jayasundara et al. (2025); Novello et al. (2025); Yeom et al. (2025). Other studies employ adaptive wavelet-positional encoding, parallel multi-scale sine responses, and frequency-adaptive audio representations Zhao et al. (2025); Han et al. (2026); Li et al. (2025). These methods establish that encoding bandwidth, activation frequency, ω_0, and weight magnitude strongly affect INR fitting. However, they primarily modify a global encoding, activation, or parameter initialisation. They do not investigate axis-dependent coordinate rescaling as a mechanism that jointly changes hidden-layer frequency atoms, target-aligned tangent dynamics, and the depth required.

Parallel and adaptively fused INRs. Multiscale and parallel designs distribute representation capacity more explicitly. Modulated periodic networks use a separate modulation pathway and local latent codes Mehta et al. (2021); BACON provides band-limited outputs at multiple levels of detail Lindell et al. (2022); MINER fits Laplacian-pyramid scales using collections of local MLPs Saragadam et al. (2022); and Neural Fourier Filter Bank combines spatial grids with frequency-specific encodings Wu et al. (2023). Split-Layer and multi-scale sine activations increase feature diversity through parallel computations within each layer Cai et al. (2026); Han et al. (2026), whereas Neural Experts partitions the coordinate domain among learned local INR experts Ben-Shabat et al. (2024). More generally, softmax-gated mixtures of experts and attention-based neural-field conditioning provide precedents for data-dependent weighting and fusion Shazeer et al. (2017); Rebain et al. (2022). These methods, however, generally employ predefined frequency bands, external pyramids or grids, local domain partitions, or parallel channels within an activation or layer. Their branches are not defined by interpretable coordinate transformations along individual axes, while

expert routing typically partitions the domain rather than forming a dense convex combination of globally evaluated, scale-specialized predictions.

CoRe-INR instead applies diagonal coordinate-rescaling matrices to the standard SIREN formulation, producing unscaled, axis-scaled, and jointly scaled branches. We connect these transformations to changes in SIREN frequency atoms, tangent features, and empirical NTK convergence. A spatial confidence module then forms a normalized, coordinate-wise fusion of all branch predictions, to give the final prediction.

3 METHOD

3.1 OBSERVATIONS ON COORDINATE SCALING

Rescaling the coordinates in the domain of INRs has the potential for performance improvements. It was initially employed for audio representation by Sitzmann et al. (2020), where the usual temporal coordinates $t \in [-1, 1]$ were rescaled to $[-100, 100]$ as a solution for the high sampling rate of audio signals. Building upon this groundwork, we further investigate the advantages of scaling in signal representation and solving inverse problems.

For the audio representation experiment, we used a SIREN (Sitzmann et al., 2020) with 2 layers and 256 units each, and recorded the learning curves under different coordinate scales as shown in Fig. 1 for the 7s Bach audio signal. Further, we computed the average frequency spectrum of each SIREN layer, using the output of each neuron. These average spectra provide insight into the spectral content supported by each layer and how the frequency information is expanded. According to Fig. 1, there is an expansion in the spectral content with coordinate scaling. Further, in this exercise, we obtained experimental results about the convergence behavior under coordinate scaling.

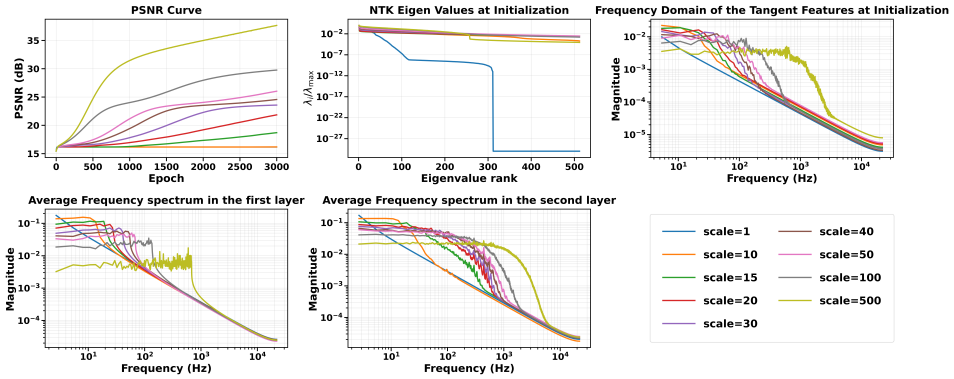


Figure 1: Observations of the scaling effect on the learning improvements in PSNR plot, Increase of the eigen values of the Neural tangent Kernel (NTK), Increase in the Tangent feature frequency domain rerpesentation, and increase in the spectral content of SIREN layers

3.2 COORDINATE SCALING FOR SPECTRAL EXPANSION

According to Fig. 1, as the scaling factor is increased, the representition of high-frequency content expands, showing that coordinate scaling has the potential to mitigate spectral bias of INRs. This spectral expansion can be theoretically explained using the derivations done by Yüce et al. (2022). Consider Equation 8 for the explanation, which shows a 2-layer INR with sinusoidal activations. Here, $\mathbf{r}$ represents the input coordinates, and $W^{(0)}$ and $W^{(1)}$ are the combination of linear layers and frequency components. The output signal is obtained through $w^{(2)}$ linear layer. .From the derivations of Yüce et al. (2022), Equation 8 can be expressed as an expansion of frequency atoms, shown in Equation 2. Accordingly, if the input coordinate $\mathbf{r}$ is replaced by a scaled version $\mathbf{ar}$, the available frequency atoms move into another scaled region of frequency, providing an expansion of frequency content for $a > 1$. In Equation 8, the bias term is not considered for simplicity. Hence, there is a difference between coordinate scaling and changing ω_0 frequency in SIREN, as it affects the bias term.

$$f_\theta(r) = \mathbf{w}^{(2)\top} \sin\Big(\mathbf{W}^{(1)} \sin\Big(\mathbf{W}^{(0)} r\Big)\Big). \tag{1}$$

$$f_\theta(r) = \sum_{m=0}^{F-1} \sum_{s_1,\ldots,s_T=-\infty}^{\infty} c_{m,s_1,\ldots,s_T} \sin\left(\left(\sum_{t=0}^{T-1} s_t \omega_t\right) r\right). \tag{2}$$

According to Equation 8, the first SIREN layer produces a set of pure sinusoids at each neuron output. These sinusoids are linearly combined by the next linear layer, which does not generate new frequency components. However, after these linear combinations are given to sine activations in the next layer, new frequency components start to emerge, resulting in the expansion of spectral components as in Fig. 1.

This frequency expansion is intuitive when it is considered from a signal processing perspective. Consider the output of the mth neuron in the second SIREN layer, shown in Equation3. The complex exponential sinusoids are similar to Frequency Modulated (FM) signals as explained by Yüce et al. (2022). These multiplications of FM signals in the time domain would become convolutions in the spectral domain. In the frequency domain, each FM type term consists of a set of impulses placed at a constant frequency interval ω_t as integer multiples $s_t\omega_t$. When these impulses are convolved, the frequency expansion occurs, while the amplitudes are determined by the combinations of the Bessel function coefficients. Further explanations and visualizations of this expansion are available in Appendix A.1.

$$z_m^{(1)} = \mathrm{Im}\left\{\prod_{t=0}^{T-1} \sum_{s_t=-\infty}^{\infty} J_{s_t}\Big(W_{m,t}^{(1)}\Big) \exp(j s_t \omega_t a r)\right\}. \tag{3}$$

However, there is spectral jitter observed as a trade-off with coordinate scaling, shown in Fig.1. This might be caused by the exposure of the microstructure of the discrete convolution as the scaling would dilate the frequency atoms, reducing the overlap within fixed FFT bins.

3.3 Coordinate Scaling for Convergence Improvement

According to the PSNR curve given in Fig. 1, for scaling factors 1 and 10, the PSNR values remain almost constant, implying the INRs are not capable of reconstructing the target audio signal. For scaling values 15 to 50, there is a certain number of epochs during which the PSNR remains constant without learning the signal, implying a behaviour of delayed learning. The delay reduces as the scaling factor increases, and for higher scales, there are no delays observable. We explain the behavior by analysing the behavior of the spectral representation of the Tangent Features (Baratin et al., 2021). Tangent Feature given by $\phi_p(\boldsymbol{x}) = \partial f_{\boldsymbol{\theta}}(\boldsymbol{x})/\partial\theta_p$, explains how each parameter adaptation would affect the learning of the function at different coordinates. We argue that the spectral representation $\widehat{\phi}_p(\boldsymbol{\omega})$ should provide insight about how the parameter adjustment would affect the learning of frequency components. For analysis, we we define the aggregate tangent spectrum given by Equation 4 where p represents all the parameters.

$$\mathcal{A}(\boldsymbol{\omega}) = \sum_p \left|\widehat{\phi}_p(\boldsymbol{\omega})\right|^2. \tag{4}$$

Since $\widehat{\phi}_p(\boldsymbol{\omega})$ relates to the sensitivity of parameters to each frequency component, a higher $\mathcal{A}(\boldsymbol{\omega})$ would reflect parameter adjustments that highly support the relevant frequency. Hence, when the target signal frequency is high, while the initialized INR has a smaller spectral representation, it needs more iterations to reach high frequencies. Through experimental results, we show that as the scaling increases, the sensitivity to high frequencies increases. Thus, it takes less number of iterations to reach the target spectrum as the initial spectral learnability is high. This is demonstrated in A.3.As the tangent features' spectral components increase, the delayed learning decreases.

Further, the convergence of an INR can be explained using the Neural Tangent Kernel (NTK), introduced by Jacot et al. (2020). The empirical NTK can be calculated as expressed in Equation 5.

$$K(\boldsymbol{x}, \boldsymbol{x}') = \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(\boldsymbol{x})^{\top} \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(\boldsymbol{x}'). \tag{5}$$

According to Tancik et al. (2020), the eigenvalues of the NTK affect the convergence of the INR to learn high-frequency details. For high eigenvalues, the convergence is improved, while the lowest eigenvalue determines the overall convergence of the INR, as they slower the convergence. As illustrated in Fig.1, as the scaling factor increases, the eigenvalues increase, including the smallest eigenvalue. Thus, through scaling, convergence can be improved. Further explanations of the NTK are available in Appendix A.2.

3.4 Coordinate Scaling for Shallow Network Performance

According to Equation 3, frequency expansion occurs through nonlinear activations as the network becomes deeper. Sitzmann et al. (2020) states that the growth of new frequencies across layers is slow. This observation applies to the unscaled case (a = 1), whereas in the scaled case (a ¿ 1), spectral expansion can be increased. Therefore, a shallow scaled network can provide better high-frequency representation than a deeper unscaled network. We experimentally validate this using SIRENs with 3, 5, and 10 hidden layers, against the scaling. In audio signal representation tasks, the shallow scaled network outperforms the deeper unscaled networks, and the ablation results are available in Appendix B.2. Thus, coordinate scaling can reduce the model size while improving high frequency representation. However, according to Figure 1, there is a reduction in the frequency coefficient as the spectral content is increased.

3.5 CoRe-INR Architecture

According to Section 3.2 and Section 3.3 coordinate scaling improves an INR in terms of its frequency representation ability and the speed of convergence. As a result, it enables shallow networks to represent high-frequency components efficiently. Building on this argument, we propose CoRe-INR, which applies coordinate scaling along different signal axes. In general, for an N-dimensional signal, the architecture consists of 2^N parallel networks. Each parallel network receives an axis-wise scaled coordinate grid as input.

Fig. 2 illustrates the overall network architecture for a two-dimensional signal. The ablation study presented in Appendix B.3 shows that there is nearly a +5 dB improvement in PSNR when coordinate scaling is introduced into an unscaled parallel network, proving that the improvement has not resulted solely from the parameter increase resulting from incorporating parallel networks, and it is mainly due to the spectral support provided by the coordinate scaling.

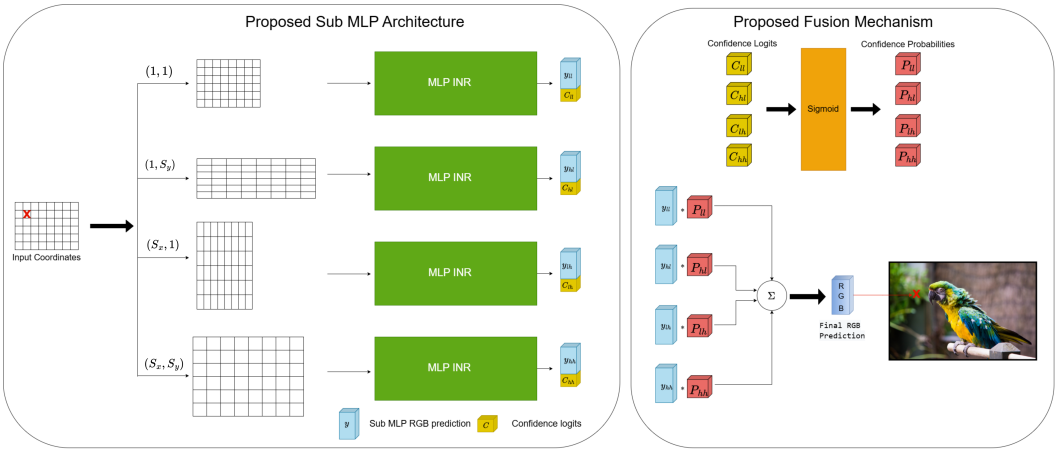


Figure 2: Architecture of CoRe-INR for 2D Signals.

Confidence-based fusion. To combine the outputs of parallel networks, we propose a confidence-based additive fusion mechanism. Let $\mathcal{S} = \{(1,1), (s_x, 1), (1, s_y), (s_x, s_y)\}$ be the coordinate scaling configurations for each branch input. For a given coordinate scaling configuration, a branch

predicts an RGB image representation $\hat{\mathbf{I}}_{\mathbf{s}}(x, y) \in \mathbb{R}^3$, and a confidence logit $c_s(x, y) \in \mathbb{R}$. The logits are converted into normalised, spatially varying weights using:

$$P_s(x, y) = \frac{\exp(c_s(x, y))}{\sum_{q \in \mathcal{S}} \exp(c_q(x, y))} \tag{6}$$

Hence, the final signal representation is given by:

$$\hat{\mathbf{I}} = \sum_{s \in \mathcal{S}} P_s(x, y) \hat{\mathbf{I}}_{\mathbf{s}}(x, y) \tag{7}$$

Learnable softmax weights enable the architecture to prioritise the branches with appropriate input scaling to represent the features present at a given local spatial location. For example, since $\mathcal{B}_{(1,1)}$ has the spectral support to represent smooth regions, the corresponding weight of $\hat{\mathbf{I}}_{(\mathbf{1},\mathbf{1})}(x, y)$ will be higher when compared to the other weights. Conversely, for spatial locations that contain directional edges and textures, corresponding branches will have higher weights. Moreover, as the non-negativity and sum-to-one constraints are imposed on the weights by the softmax operation, the final signal reconstruction lies within a stable convex combination, precluding uncontrolled amplification of the branch outputs.

4 Experiments & Results

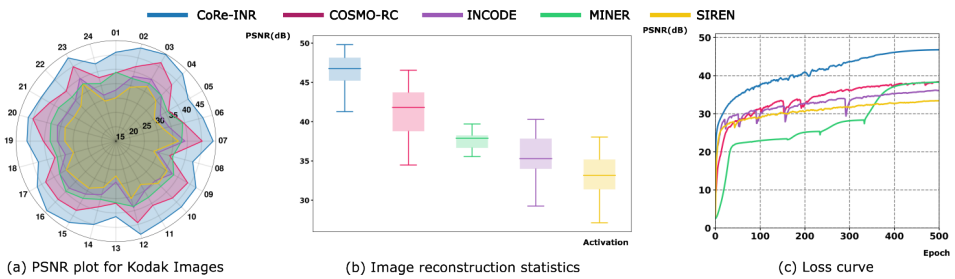
We implement all models in the **PyTorch framework** and train them with the Adam optimizer on an **NVIDIA RTX 6000 ADA** GPU with 48 GB of memory. Image tasks use four parallel three-layer MLP branches of width 256, while occupancy and audio reconstruction tasks use eight and two branches, respectively. The fusion temperature is annealed from 1.0 to promote deterministic confidence-based selection. Additional results and task-specific settings are provided in Appendix C, and computational cost, runtime, throughput, and accuracy trade-offs are analyzed in Appendix D.

4.1 Image Representation

Data. We conducted our experiments on the Kodak (Franzen, n.d.) dataset, with 24 lossless images. Each model was trained on the images' original resolution.

Implementation Details. The input image is downscaled by $1/2$. The model is trained for 500 epochs with a batch size of 256×256, a learning rate of 1.5×10^{-4}, and a decay factor of 0.1, requiring approximately 3.7 GB of GPU memory. The frequency scaling factor was set to 0.5.

Results. As shown in Fig. 3, CoRe-INR outperforms every SOTA method on the Kodak dataset for all images, achieving an average PSNR of 46.4 dB alongside faster, more stable convergence. Qualitative reconstruction results are provided in Appendix C.1.



(a) PSNR plot for Kodak Images (b) Image reconstruction statistics (c) Loss curve

Figure 3: Quantitative image reconstruction results produced by CoRe-INR.

4.2 Image Denoising

Data. To evaluate robustness against noisy input signals, we test on the *Parrot* image from the DIV2K dataset (Timofte et al., 2018).

Implementation Details. The input image is downscaled by $1/2$ and noised with a photon noise of 4×10^1 and additive noise with an SNR of 2, yielding a PSNR of 16.73 dB. Models are trained for 500 epochs with a batch size of 256×256, a learning rate of 1.5×10^{-4}, and a decay factor of 0.1, requiring approximately 3.7 GB of GPU memory. A frequency scaling factor of 0.5 suppresses high-frequency noise while preserving the image's global structure.

Results. Reconstruction quality is evaluated using the PSNR metric against the ground-truth images. As shown in Fig. 4, CoRe-INR achieves a PSNR of 30.90 dB, outperforming COSMO-RC while maintaining superior structural fidelity. Notably, CoRe-INR preserves fine anatomical details in challenging regions such as the beak, where competing methods tend to oversmooth or introduce abnormal artifacts. This indicates improved robustness to high-frequency noise and better edge preservation under the evaluated conditions and noise configurations. Additional qualitative results are shown in the Appendix C.3.

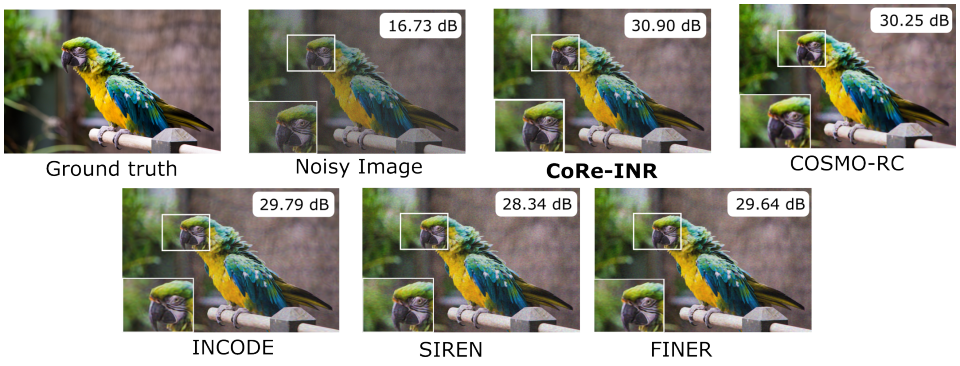


Figure 4: Qualitative denoising results comparing CoRe-INR against existing approaches.

4.3 Image Super-Resolution

Data. We use a high-resolution image from the DIV2K dataset with resolution $1356 \times 2040 \times 3$.

Implementation Details. We transform the sample high-resolution image as follows for evaluation purposes. Low-resolution inputs are generated using downsampling factors of $1/2$, $1/4$, and $1/6$, corresponding to $2\times$, $4\times$, and $6\times$ super-resolution tasks. Models are trained for 500 epochs using a batch size of 256×256, learning rate 9×10^{-4}, and decay factor 0.1. The architecture requires approximately 2,250 MB GPU memory. Frequency scaling factors of 0.3, 0.5, and 0.9 are used for $2\times$, $4\times$, and $6\times$ settings respectively.

Results. Scaling quality is evaluated using the PSNR metric against the ground-truth images. Table 1 shows quantitative results, where CoRe-INR achieves the highest PSNR at all three scales and the highest or tied-highest SSIM. The largest PSNR gain over COSMO-RC is 0.40 dB at $2\times$. Additional qualitative results are given in the Appendix C.2.

Table 1: Comparison of super-resolution performance with existing state-of-the-art INR methods. The best result for each metric is highlighted in bold.

METHOD	2×		4×		6×	
	PSNR (dB) ↑	SSIM ↑	PSNR (dB) ↑	SSIM ↑	PSNR (dB) ↑	SSIM ↑
ReLU+PE	32.80	0.91	28.89	0.87	26.29	0.83
SIREN	32.26	0.90	29.62	0.87	27.31	0.81
INCODE	32.83	0.90	29.96	0.85	26.63	0.78
FINER	32.94	0.91	29.75	0.84	27.02	0.80
COSMO-RC	34.03	0.96	30.42	**0.95**	27.66	0.83
CoRe-INR	**34.43**	**0.97**	**30.48**	**0.95**	**27.68**	**0.85**

4.4 IMAGE INPAINTING

Data. We use the *Celtic Spiral Knots* image with resolution $572 \times 582 \times 3$.

Implementation Details. A random mask is applied such that only 20% of pixel coordinates of the image are observed during training. Evaluation is performed by reconstructing the full image grid and computing the PSNR against the ground-truth images. Training uses 500 epochs with batch size 256×256, learning rate 1.5×10^{-4}, and decay factor 0.25. The model requires approximately 700 MB GPU memory. A frequency scaling factor of 0.5 is used as described in Section 3.

Results. As shown in the Fig. 5, CoRe-INR achieves a reconstruction quality of 22.07 dB PSNR, outperforming COSMO-RC (21.88 dB). The zoomed view highlights superior structural smoothness and more natural pixel fidelity over other methods.

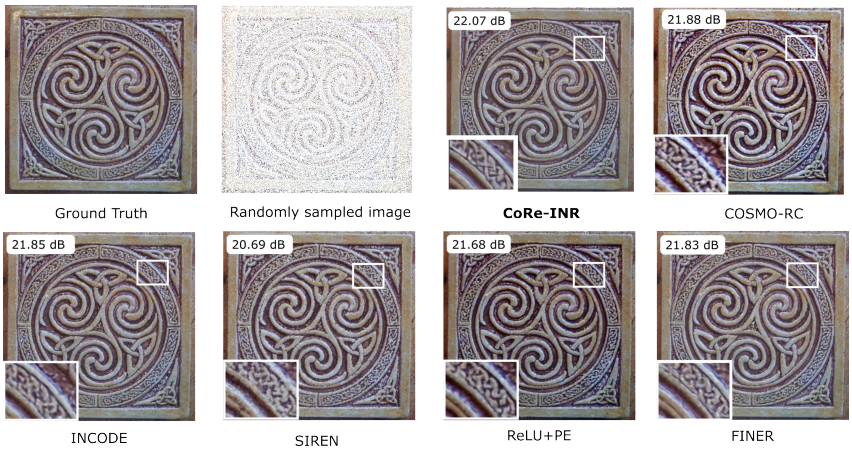


Figure 5: Qualitative image inpainting results produced by CoRe-INR.

4.5 3D OCCUPANCY

Data. We use the **Lucy** dataset and convert meshes into a 512^3 occupancy grid with binary labels indicating occupied and empty voxels.

Implementation Details. Unlike other tasks, here for the occupancy task $N = 2^3$ parallel pipelines are used to model the full combinations of coordinate scaling across three spatial axes. We train the model for 200 epochs using 10^6 coordinate samples per batch, a learning rate of 1×10^{-4}, and a decay factor of 0.2. Due to volumetric complexity, training requires approximately 20 GB GPU memory.

Results. Fig. 6 shows that CoRe-INR achieves the highest Intersection over Union (IoU) of 0.999, outperforming COSMO-RC (0.995), FINER (0.994), INCODE (0.993), and SIREN (0.992). Although visual differences are subtle to the human eye, the IoU gains confirm our method's superior reconstruction performance.

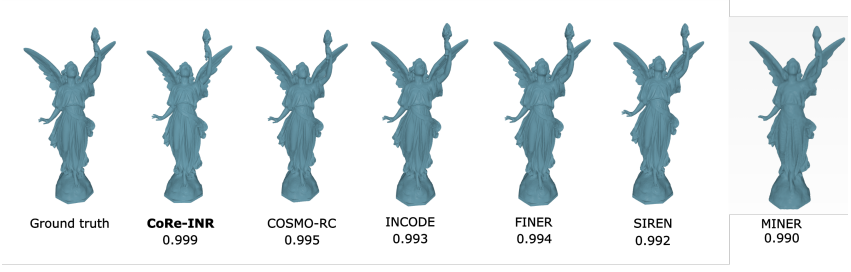


Figure 6: 3D occupancy reconstruction results obtained using CoRe-INR.

4.6 AUDIO RECONSTRUCTION

Data. We use the first 7 seconds of Bach's Cello Suite No. 1 to evaluate audio reconstruction performance in terms of PSNR.

Implementation Details. CoRe-INR employs two parallel MLP branches modeling low and high frequency components. Frequency scaling is set to 100 for low-frequency and 400 for high-frequency components following prior work. The model uses 256 hidden dimensions per branch with 3 SIREN layers and is trained for 1000 epochs for fair comparison with INCODE.

Results. CoRe-INR achieves 50.02 dB, outperforming the strongest baseline, INCODE, by 0.92 dB (Table 2). This demonstrates strong effectiveness in modeling high-fidelity temporal signals.

Table 2: Comparison of INR methods for audio-signal representation. The best result is highlighted in bold.

METHOD	PSNR (dB) ↑
SIREN	37.92
INCODE	49.10
Gaussian	38.50
ReLU+PE	22.99
CoRe-INR	**50.02**

4.7 LIMITATIONS

The main focus of the research is to provide theoretical explanations for the potential use of coordinate scaling for implicit neural representations. Here, we showed that the tangent features can be useful to explain the convergence properties of the INR, in addition to the usual NTK-based explanations. Further, we demonstrated that scaling along the coordinate axes can improve high-dimensional signal representation tasks by formulating the CoRe-INR architecture. However, if the scaling is set to extremely high values, representation performance will drop due to the traed-off. Hence, there is an optimal scaling range that might depend on the signal characteristics. Hence, the proposed tangent feature-based arguments can be extended to formulate optimal scaling factors. Moreover, since the scaling is global to the entire signal, there might be a clamp in reaching the most optimal achievable frequency representation based on the local context.

5 CONCLUSION

In this paper, we explored the effect of coordinate scaling using a global scaling factor on the INRs input coordinates with both theoretical and experimental approaches. Using the frequency atoms, we showed that there is an obvious spectral support that could be improved by scaling. However, extreme scaling would create spectral jitters, showing that there is a trade-off for scaling. Further, we observed that there are situations where the INRs get delayed in starting to learn the signal, as it takes multiple iterations to change the parameters to increase the spectral support. Using a novel approach with the spectral representation of tangent features, we show that by scaling, the spectral sensitivity of parameters can be improved, resultsing in less amout of iterations to reach the target signal. Moreover, we propose CoRe-INR that shows the significance of directional scaling, where the coordinates are scaled along multiple axis of high-dimensional signals. CoRe-INR's directional scaling reveals that it is capable in representing high dimensional tasks and solving inverse problems. We demonstrated the capability of CoRe-INR in representation tasks for audio, image and 3D objects and inverse problems on denoising, super resolution and inpainting, where the SOTA algorithms are beaten. The foundation that we explored for coordinate rescaling opens a new avenue to explore the scaling in different aspects such as locally varying coordinate scaling.

AI USE STATEMENT

In this work, we used generative AI tools for grammar checking and correction, LaTeX code generation, and identification of relevant literature. We did not use generative AI tools for research

tasks, including the development of theoretical models or conceptual frameworks, formulation of mathematical claims, writing of mathematical proofs, refinement of hypotheses, design of research methodology, implementation of methods. Other required-disclosure tasks, including synthetic dataset generation, translation, dataset cleaning or reformatting, and qualitative or thematic data analysis, were not applicable to this work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

The authors do not have concerns about code of ethics.

Reproducibility statement

The proposed method is based on the existing SIREN activation, where its performance improvements are done using coordinate scaling. For the CoRe-INR architecture, the implementation details of the parallel branches and fusion mechanism are explained in the model architecture.

References

Aristide Baratin, Thomas George, César Laurent, R Devon Hjelm, Guillaume Lajoie, Pascal Vincent, and Simon Lacoste-Julien. Implicit regularization via neural feature alignment. In Arindam Banerjee and Kenji Fukumizu (eds.), *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2269–2277. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/baratin21a.html.

Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, Sameera Ramasinghe, and Stephen Gould. Neural experts: Mixture of experts for implicit neural representations. In *Advances in Neural Information Processing Systems*, volume 37, pp. 101641–101670, 2024. doi: 10.52202/079017-3224.

Zhicheng Cai, Hao Zhu, Linsen Chen, Qiu Shen, and Xun Cao. Split-layer: Enhancing implicit neural representation by maximizing the dimensionality of feature space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 2561–2570, 2026. doi: 10.1609/aaai.v40i4.37243.

Richard W. Franzen. True color kodak images, n.d. Accessed: 2026-07-29.

Jufeng Han, Shu Wei, Min Wu, Lina Yu, Weijun Li, Linjun Sun, Hong Qin, and Yan Pang. Implicit neural representation with multi-scale sine activation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 21567–21575, 2026. doi: 10.1609/aaai.v40i26.39305.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks, 2020. URL https://arxiv.org/abs/1806.07572.

Dhananjaya Jayasundara, Heng Zhao, Demetrio Labate, and Vishal M. Patel. Mire: Matched implicit neural representations. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8279–8288, 2025. doi: 10.1109/CVPR52734.2025.00775.

Amirhossein Kazerouni, Reza Azad, Alireza Hosseini, Dorit Merhof, and Ulas Bagci. Incode: Implicit neural conditioning with prior knowledge embeddings. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1287–1296, 2024. doi: 10.1109/WACV57701.2024.00133.

Linfei Li, Lin Zhang, Zhong Wang, Fengyi Zhang, Zelin Li, and Ying Shen. Representing sounds as neural amplitude fields: A benchmark of coordinate-MLPs and a fourier kolmogorov–arnold framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 24458–24466, 2025. doi: 10.1609/aaai.v39i23.34624.

David B. Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation, 2022. URL https://arxiv.org/abs/2112.04645.

Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao. Finer: Flexible spectral-bias tuning in implicit neural representation by variableperiodic activation functions. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2713–2722, 2024. doi: 10.1109/CVPR52733.2024.00262.

Ishit Mehta, Michaël Gharbi, Connelly Barnes, Eli Shechtman, Ravi Ramamoorthi, and Manmohan Chandraker. Modulated periodic activations for generalizable local functional representations, 2021. URL https://arxiv.org/abs/2104.03960.

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4455–4465, 2019. doi: 10.1109/CVPR.2019.00459.

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. *Commun. ACM*, 65(1):99–106, December 2021. ISSN 0001-0782. doi: 10.1145/3503250. URL https://doi.org/10.1145/3503250.

Tiago Novello, Diana Aldana, Andre Araujo, and Luiz Velho. Tuning the frequencies: Robust training for sinusoidal neural networks. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3071–3080, 2025. doi: 10.1109/CVPR52734.2025.00292.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5301–5310. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/rahaman19a.html.

Sameera Ramasinghe and Simon Lucey. Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (eds.), *Computer Vision – ECCV 2022*, pp. 142–158, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-19827-4.

Daniel Rebain, Mark J. Matthews, Kwang Moo Yi, Gopal Sharma, Dmitry Lagun, and Andrea Tagliasacchi. Attention beats concatenation for conditioning neural fields, 2022. URL https://arxiv.org/abs/2209.10684.

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

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G. Baraniuk. Wire: Wavelet implicit neural representations. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18507–18516, 2023. doi: 10.1109/CVPR52729.2023.01775.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=B1ckMDqlg.

Kexuan Shi, Xingyu Zhou, and Shuhang Gu. Improved Implicit Neural Representation with Fourier Reparameterized Training . In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25985–25994, Los Alamitos, CA, USA, June 2024. IEEE Computer Society. doi: 10.1109/CVPR52733.2024.02455. URL https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.02455.

Vincent Sitzmann, Michael Zollhoefer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/b5dc4e5d9b495d0196f61d45b26ef33e-Paper.pdf.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/53c04118df112c13a8c34b38343b9c10-Paper.pdf.

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. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Pandula Thennakoon, Avishka Ranasinghe, Mario De Silva, Buwaneka Epakanda, Roshan Godaliyadda, Mervyn Parakrama Bandara Ekanayake, and Vijitha R. Herath. COSMO-INR: Complex sinusoidal modulation for implicit neural representations. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=yGJrvSU6wK.

Radu Timofte, Shuhang Gu, Jiqing Wu, Luc Van Gool, Lei Zhang, Ming-Hsuan Yang, Muhammad Haris, et al. Ntire 2018 challenge on single image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

Kushal Vyas, Ashok Veeraraghavan, and Guha Balakrishnan. Fit Pixels, Get Labels: Meta-Learned Implicit Networks for Image Segmentation . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume LNCS 15962. Springer Nature Switzerland, September 2025.

Zhijie Wu, Yuhe Jin, and Kwang Moo Yi. Neural fourier filter bank. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14153–14163, 2023. doi: 10.1109/CVPR52729.2023.01360.

Taesun Yeom, Sangyoon Lee, and Jaeho Lee. Fast training of sinusoidal neural fields via scaling initialization. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Sr5XaZzirA.

Gizem Yüce, Guillermo Ortiz-Jiménez, Beril Besbinar, and Pascal Frossard. A structured dictionary perspective on implicit neural representations. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19206–19216, 2022. doi: 10.1109/CVPR52688.2022.01863.

Hongxu Zhao, Zelin Gao, Yue Wang, Rong Xiong, and Yu Zhang. Adaptive wavelet-positional encoding for high-frequency information learning in implicit neural representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10430–10438, 2025. doi: 10.1609/aaai.v39i10.33132.

Yushen Zuo, Qi Zheng, Mingyang Wu, Xinrui Jiang, Renjie Li, Jian Wang, Yide Zhang, Gengchen Mai, Lihong V. Wang, James Zou, Xiaoyu Wang, Ming-Hsuan Yang, and Zhengzhong Tu. 4kagent: Agentic any image to 4k super-resolution. 2025. URL https://arxiv.org/abs/2507.07105.

A Further Theoretical Explanation on Scaling

In this section, we theoretically and experimentally validate and further explain the coordinate scaling in INRs.

A.1 Two Layer SIREN and the FM Convolution

The mathematical explanations in this section is based on the proof in (Yüce et al., 2022). In the explanations, the bias term is not considered, as it only changes the phase and does not affect coordinate scaling (But when changing ω_0 frequencies, the bias term also largely changes)

Consider an INR that has 2 sinusoidal activated layers, which create a mapping function,

$$f_\theta(r) = \mathbf{w}^{(2)\top} \sin\Big(\omega_1 \mathbf{W}^{(1)} \sin\Big(\omega_0 \mathbf{W}^{(0)} r\Big)\Big). \tag{8}$$

where, $\mathbf{W}^{(0)} \in \mathbb{R}^T$, $\mathbf{W}^{(1)} \in \mathbb{R}^{T\times T}$, $\mathbf{W}^{(2)} \in \mathbb{R}^T$ and $r \in \mathbb{R}$. Let, $\mathbf{z^{(0)}} = \sin(\omega_0 \mathbf{W^{(0)}} \mathbf{r})$. Thus, the first layer outputs a set of pure sinusoids, with frequencies $(\omega_0 \mathbf{W}^{(0)})$. In the next linear layer $(\mathbf{W^{(1)}} \mathbf{z^{(0)}})$, these pure sinusoids are linearly combined, thus no new frequency content is generated. When these frequency combinations are applied to the next $\sin$ activation non-linearity, new frequency content starts to emerge.

Output at the second activation, $z^{(1)} = \sin(\omega_1 \mathbf{W^{(1)}} \mathbf{z^{(0)}})$ creates new frequency content from the $\sin$ non-linearity. Consider the $\mathrm{m^{th}}$ neuron of the output.

$$\begin{aligned}
z_m^{(1)} &= \sin\left(\omega_1 \sum_{t=0}^{T-1} W_{m,t}^{(1)} \sin\Big(\omega_0 W_t^{(0)} r\Big)\right) \\
&= \mathrm{Im}\left\{\exp\left(\mathrm{j}\omega_1 \sum_{t=0}^{T-1} W_{m,t}^{(1)} \sin\Big(\omega_0 W_t^{(0)} r\Big)\right)\right\} \\
&= \mathrm{Im}\left\{\prod_{t=0}^{T-1} \exp\Big(\mathrm{j}\omega_1 W_{m,t}^{(1)} \sin\Big(\omega_0 W_t^{(0)} r\Big)\Big)\right\} \\
&= \mathrm{Im}\left\{\prod_{t=0}^{T-1} \sum_{s_t=-\infty}^{\infty} J_{s_t}\Big(\omega_1 W_{m,t}^{(1)}\Big) \exp\Big(\mathrm{j} s_t \omega_0 W_t^{(0)} r\Big)\right\}
\end{aligned} \tag{9}$$

If we consider Equation 9 in the frequency domain, this multiplication of FM-type signals becomes a convolution. We demonstrate the expansion through this convolution in Fig. 7, using a layer with 5 neurons and the Bessel order $S_t = 2$. The frequency expansion of this tiny network can be used to imagine how a large-width neural network would expand the frequency content by the sine non-linearity. Further, by replacing the input coordinate r with coordinate scaling ar, where a is a scalar, we can expand the frequency content towards higher-frequency support as shown in Fig. 8.

A.2 Convergence of INRs with Coordinate Rescaling

Consider an INR $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ evaluated over the training coordinates, with predictions $\boldsymbol{f}_{\boldsymbol{\theta}(t)}$ and target values $\boldsymbol{y}$. The residual vector is defined as

$$\boldsymbol{e}(t) = \boldsymbol{f}_{\boldsymbol{\theta}(t)} - \boldsymbol{y}. \tag{10}$$

For the squared-error objective,

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2} \left\|\boldsymbol{e}(t)\right\|_2^2, \tag{11}$$

gradient-flow training evolves the parameters according to

$$\frac{d\boldsymbol{\theta}(t)}{dt} = -\eta \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}(t)), \tag{12}$$

where η denotes the learning rate.

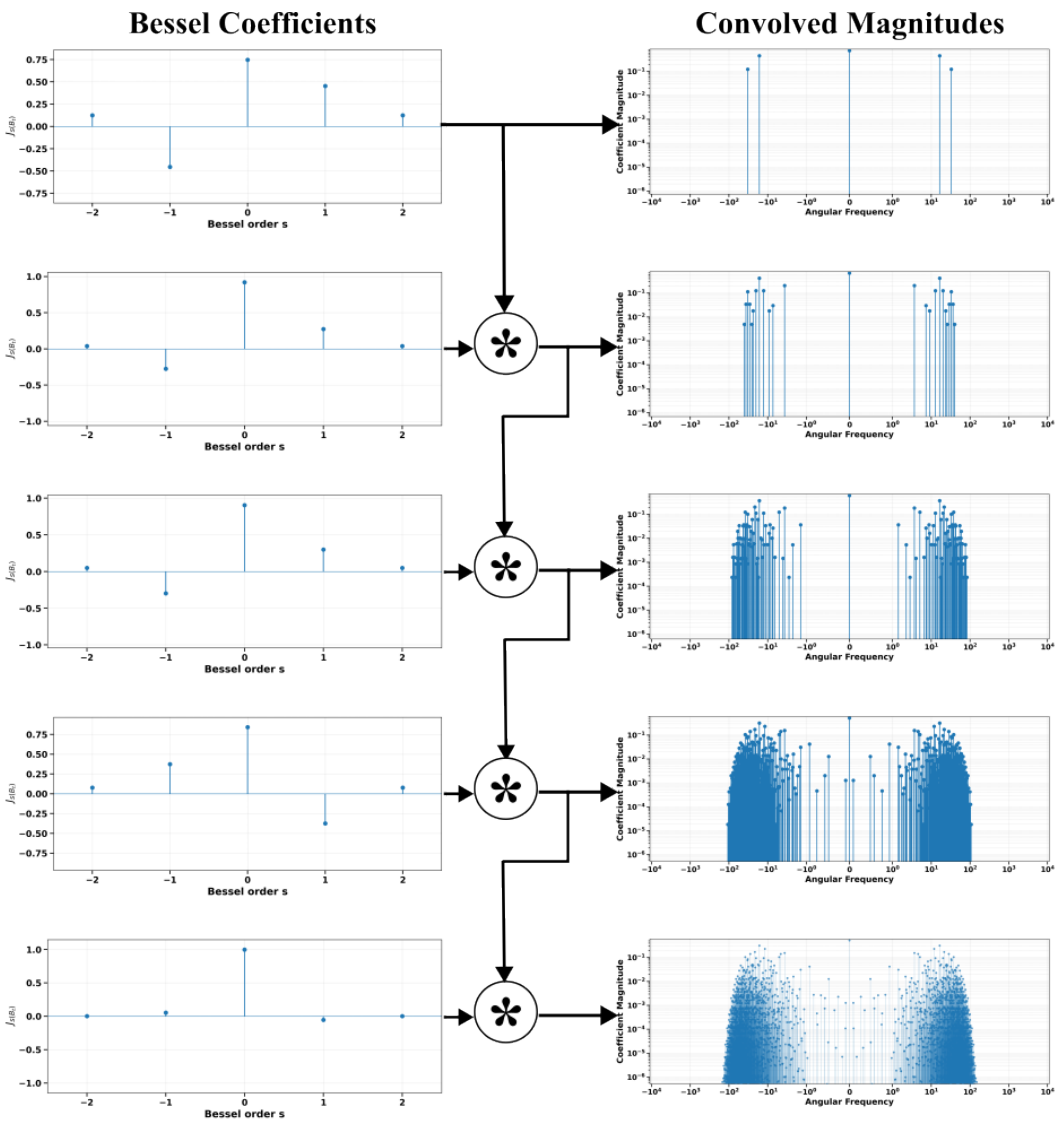


Figure 7: The convolution of the FM-type signals and the expansion of the frequency content for a tiny-width layer of 5 neurons ($*$ represents the convolution operator)

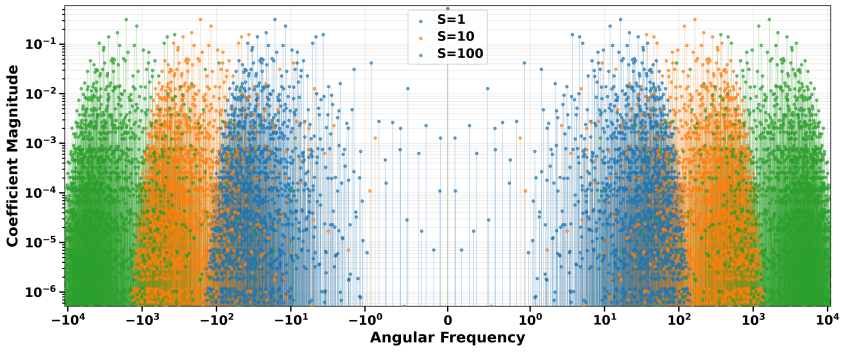


Figure 8: Convolved result for different scaling factors (1,10 and 100)

Let $\mathbf{J}_{\boldsymbol{\theta}(t)}$ denote the Jacobian of the network outputs with respect to the parameters. Using

$$\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}(t)) = \mathbf{J}_{\boldsymbol{\theta}(t)}^{\top}\boldsymbol{e}(t), \tag{13}$$

the parameter dynamics become

$$\frac{d\boldsymbol{\theta}(t)}{dt} = -\eta\mathbf{J}_{\boldsymbol{\theta}(t)}^{\top}\boldsymbol{e}(t). \tag{14}$$

The corresponding evolution of the network predictions is

$$\frac{d\boldsymbol{f}_{\boldsymbol{\theta}(t)}}{dt} = \mathbf{J}_{\boldsymbol{\theta}(t)}\frac{d\boldsymbol{\theta}(t)}{dt} \tag{15}$$

$$= -\eta\mathbf{J}_{\boldsymbol{\theta}(t)}\mathbf{J}_{\boldsymbol{\theta}(t)}^{\top}\boldsymbol{e}(t). \tag{16}$$

The empirical NTK matrix is

$$\mathbf{K}(t) = \mathbf{J}_{\boldsymbol{\theta}(t)}\mathbf{J}_{\boldsymbol{\theta}(t)}^{\top}, \tag{17}$$

whose entries correspond to

$$K(\boldsymbol{x}, \boldsymbol{x}') = \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(\boldsymbol{x})^{\top}\nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(\boldsymbol{x}'). \tag{18}$$

Since the target $\boldsymbol{y}$ is constant, the residual dynamics satisfy

$$\frac{d\boldsymbol{e}(t)}{dt} = -\eta\mathbf{K}(t)\boldsymbol{e}(t) \tag{19}$$

Under the linearised NTK regime, the kernel remains approximately constant, $\mathbf{K}(t) \approx \mathbf{K}(0) = \mathbf{K}$. Therefore,

$$\boldsymbol{e}(t) = \exp(-\eta\mathbf{K}t)\,\boldsymbol{e}(0). \tag{20}$$

Using the eigendecomposition

$$\mathbf{K} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\top}, \qquad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N), \tag{21}$$

the residual can be expressed as

$$\boldsymbol{e}(t) = \sum_{i=1}^{N} \alpha_i(0)\exp(-\eta\lambda_i t)\mathbf{u}_i, \tag{22}$$

where

$$\alpha_i(t) = \mathbf{u}_i^{\top}\boldsymbol{e}(t). \tag{23}$$

Hence, each residual component evolves as

$$\alpha_i(t) = \alpha_i(0)\exp(-\eta\lambda_i t) \tag{24}$$

The corresponding training loss is

$$\mathcal{L}(t) = \frac{1}{2}\sum_{i=1}^{N} \alpha_i^2(0)\exp(-2\eta\lambda_i t). \tag{25}$$

Therefore, the convergence rate of each residual mode is controlled by its associated NTK eigenvalue λ_i. Modes associated with larger target-relevant eigenvalues decay more rapidly, whereas modes associated with small eigenvalues converge slowly. In our experiments, suitable scaling increases the eigenvalues associated with target-relevant signal components, resulting in faster residual contraction.

A.3 Tangent Features Interpretation for Delayed Learning

Tangent features in the coordinate domain. For an INR $f_{\boldsymbol{\theta}}(\boldsymbol{x})$, the tangent feature associated with parameter θ_p is defined as

$$\phi_p(\boldsymbol{x}) = \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \theta_p}. \tag{26}$$

It describes how strongly a small change in θ_p modifies the network output at coordinate $\boldsymbol{x}$. For a small parameter perturbation $\Delta\boldsymbol{\theta}$, the corresponding first-order change in the network output is

$$\Delta f(\boldsymbol{x}) \approx \sum_p \phi_p(\boldsymbol{x}) \Delta\theta_p. \tag{27}$$

Therefore, the value of $\phi_p(\boldsymbol{x})$ can vary across the signal domain: changing the same parameter may strongly affect the output at some coordinates while having a weaker effect at others. Collectively, the tangent features describe the local set of functional changes that can be produced by small parameter updates at the current stage of training.

Frequency-domain interpretation. Taking the Fourier transform of Equation 27 gives

$$\Delta \widehat{f}(\boldsymbol{\omega}) \approx \sum_p \widehat{\phi}_p(\boldsymbol{\omega}) \Delta\theta_p, \tag{28}$$

where

$$\widehat{\phi}_p(\boldsymbol{\omega}) = \mathcal{F}\{\phi_p(\boldsymbol{x})\}(\boldsymbol{\omega}). \tag{29}$$

While $\phi_p(\boldsymbol{x})$ measures how a parameter change affects the output at a particular coordinate, its Fourier transform $\widehat{\phi}_p(\boldsymbol{\omega})$ measures how strongly that parameter change can modify the output at frequency $\boldsymbol{\omega}$. A large magnitude at a particular frequency indicates that the corresponding parameter direction can efficiently alter that frequency component through a small local update. Conversely, a small magnitude indicates that the frequency is weakly accessible from the current parameter configuration.

Aggregate tangent spectrum. Since the network contains many parameters, we summarise their collective frequency sensitivity using the aggregate tangent spectrum

$$\mathcal{A}(\boldsymbol{\omega}) = \sum_p \left|\widehat{\phi}_p(\boldsymbol{\omega})\right|^2. \tag{30}$$

This quantity measures the total local sensitivity of the network output to parameter perturbations at frequency $\boldsymbol{\omega}$. A large aggregate value means that the network has many, or sufficiently strong, parameter directions capable of modifying that frequency. A small value means that the frequency is difficult to change through the currently available local parameter updates.

The aggregate tangent spectrum should not be interpreted as the spectrum of the network output itself. Instead, it represents the spectrum of the changes accessible to the network around its current parameters. A frequency may therefore be present in the target signal but remain weakly represented in the tangent spectrum, indicating that the network cannot yet efficiently reduce the residual at that frequency.

Tangent support and delayed fitting. At small coordinate-scaling factors, the tangent spectrum at initialization provides weak support over the high-frequency range of the target signal. Consequently, initial parameter updates cannot strongly modify these target frequencies. The network must first adjust its parameters and reshape its tangent features until sufficient high-frequency sensitivity is developed. This feature-adaptation period appears in the training curves as an initial PSNR plateau as shown in the PSNR curves of Fig. 9.

The aggregate-spectrum heat maps support this interpretation. During the initial flat region of the PSNR curve, the tangent spectrum contains little energy over the target high-frequency band. When the PSNR begins to increase, the heat maps simultaneously show stronger tangent coverage in this region. This correspondence indicates that effective fitting begins when the network develops sufficient local sensitivity to modify the missing high-frequency components.

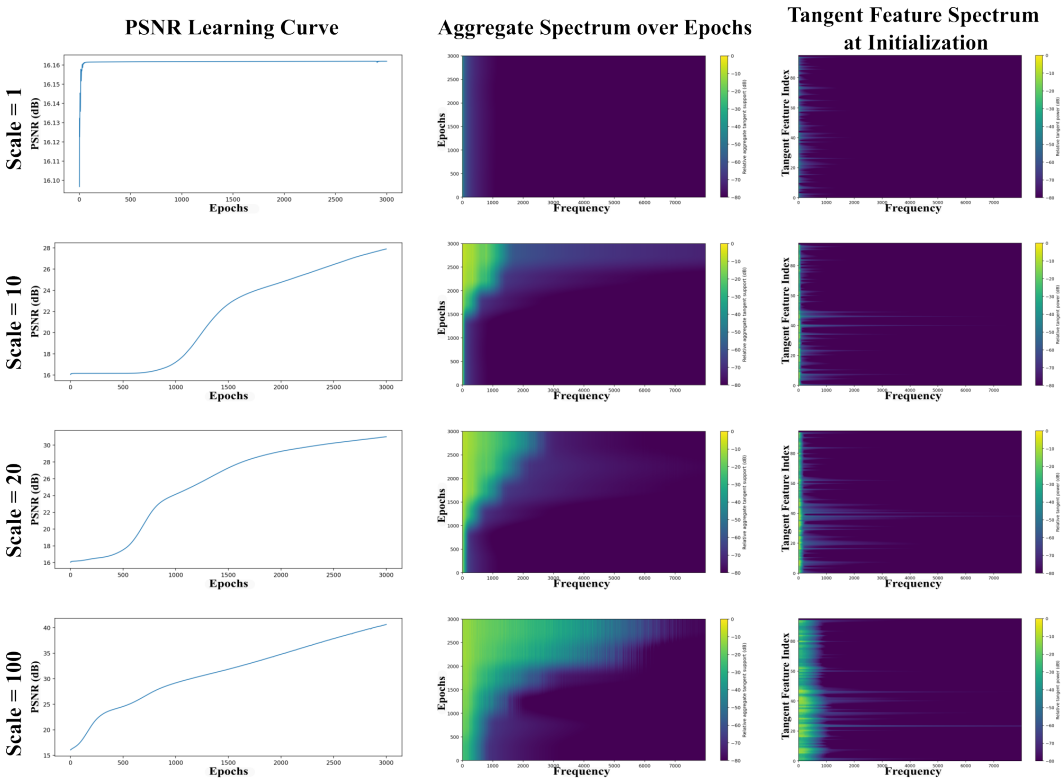


Figure 9: Learning curves (Left) under different scaling values,the aggregate spectrum variation across frequencies during training epochs (Middle), and the frequency representation of a few tangent features at initialization (Right). It is visible that the spectrum starts to move towards high-frequency content when the PSNR curve starts to increase from being flattened. As the scaling value increases, more frequency content becomes highly sensitive to parameter changes; hence, with fewer parameter changes, the signal can be easily learned.

Coordinate scaling reduces this initial spectral gap by providing more support towards high-frequency regions in the tangent features at initialization (Fig. 9 right-side spectra). Larger scaling factors shift the tangent features towards higher frequencies and provide stronger coverage of the target bandwidth at, or shortly after, initialisation. Therefore, less parameter adaptation is required before the high-frequency residual becomes accessible. The initial plateau consequently becomes shorter, and for sufficiently large scaling factors, learning may begin immediately. In contrast, smaller scaling factors require more optimisation epochs to transform the tangent geometry before meaningful high-frequency fitting can occur.

The PSNR curves and aggregate-spectrum heat maps therefore provide complementary evidence. The PSNR curves show when effective fitting begins, whereas the tangent-spectrum heat maps indicate which frequency changes become locally accessible at that time. Their observed correspondence supports the interpretation that coordinate scaling reduces the feature-adaptation burden required to reach the target bandwidth. In this sense, scaling reduces the spectral distance between the initial tangent geometry and the tangent geometry required to reconstruct the target signal.

B Ablation Studies for Coordinate Scaling

B.1 Effect of Coordinate Scaling in parallel branches of CoRe-INR

We examined the effect of scaling along different axes in CoRe-INR, with the corresponding confidence values. We demonstrate the results for the Kodak 01 image with the scaling factor of 6. According to Fig. 10, by scaling along an axis, more edge information corresponding to that direction

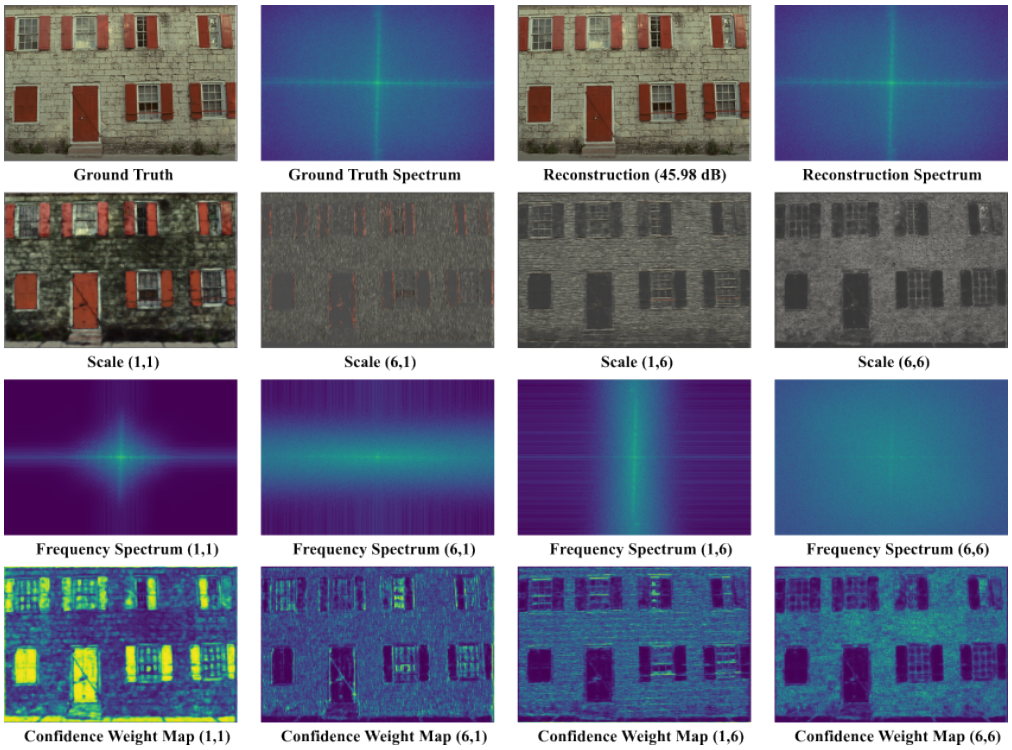


Figure 10: Scaling along different axes highlights the edge information. This is further observed through the frequency spectrum of each branch output. The corresponding confidence further supports the explanation

has been learned by the corresponding parallel branch. The spectral domain of each branch further proves that more high-frequency content is learned along the corresponding direction of scaling. For the scaling of (S_x, S_y), the diagonal frequency-related information are learned by that branch. The $(1, 1)$ branch focuses more on the smoother regions of the image. Further, the confidence maps of each parallel branch suggest that it has a high weight for the corresponding branch for the edge information along each axis when fusing the 4 branch outputs.

B.2 Spectral Expansion in Deep Networks

Table 3: Comparison of scaled shallow networks and unscaled deep networks.

MODEL	PARAMETER COUNT	PSNR (dB)
3-layer SIREN: No scaling	132353	16.16
3-layer SIREN: Scaled by 100	132353	40.65
5-layer SIREN: No scaling	263937	16.16
5-layer SIREN: Scaled by 100	263937	32.53
10-layer SIREN: No scaling	592897	16.16
10-layer SIREN: Scaled by 100	592897	36.73

We examined how the frequency expands when the INR has more hidden layers. In this experiment, we conducted the analysis for INRs having layers of 3,5 and 10 for audio signal representation for the first 7s of the *bach* audio file, training for 3000 epochs. Table 3 shows that, even when the INR is significantly deeper, it cannot learn the signal compared to a shallow INR with scaling. Further, even with scaling, a deeper network the performance is less than a shallow network for this audio representation task. According to Fig. 11, a shallow network with scaling has high frequency content compared to an unscaled deeper network. Moreover, even when scaling was used for the deeper network, it has achieved a similar maximum spectral content to the shallower network, with a

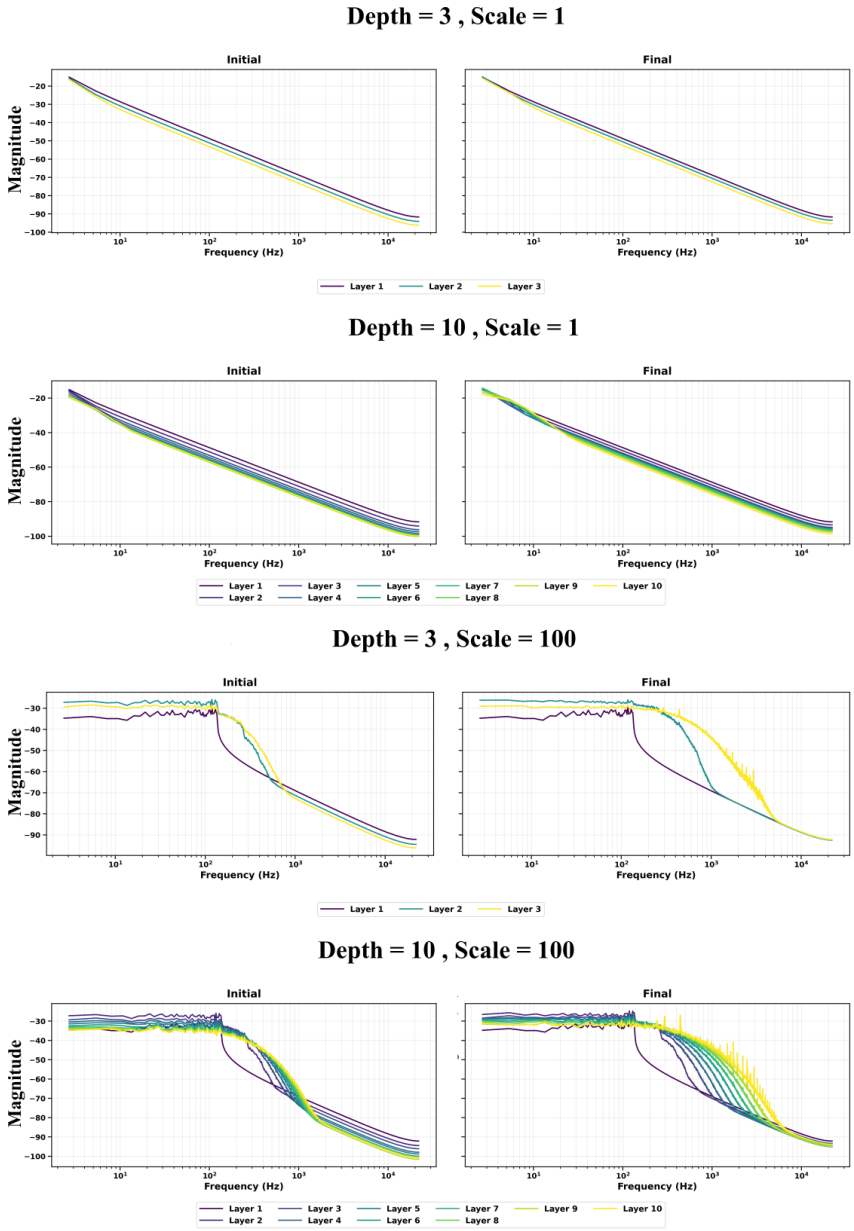


Figure 11: Effect of coordinate scaling for shallow and deep networks with and without scaling. Shallow networks can perform better compared to deeper networks when the coordinates are scaled.

slow rate of frequency content increment by each layer. Thus, there is a potential advantage of using shallower networks with scaling over deeper networks; hence, the INR representation capability can be improved with less parameter usage.

B.3 Ablation on Model Capacity and Scaling

To empirically prove that scaling has an impact on representation improvements above increasing the number of model parameters, we conducted several ablation studies for Kodak image 20 with the same training details as in Appendix C. Table 4 shows that having the same parallel branches with and without scaling has a difference of about 5 dB. Further, using the confidence-based fusion has improved the image representation task by 1.7 dB. By model parameter increments, we can improve the representation results. But on top of that, by scaling we can achieve a similar result with fewer

Table 4: Effect of model capacity and coordinate scaling on representation performance.

MODEL	SCALING FACTOR	FUSION	PARAMETERS	PSNR (dB)
Four Branches	1	Average	532492	40.79
Four Branches	6	Average	532492	45.11
Four Branches	6	Confidence	532492	46.80
Single Branch (1024 width, 3 Layers)	1	N/A	2105347	46.90
Three branches Without (S_x, S_y)	6	Confidence	400140	43.78

parameters, when comparing with the single INR of 1024 width. Further, we examined that (S_x, S_y) scaling branch is essential to improve the signal representation as it can represent frequency content along diagonal directions as shown in Figure 10.

C Additional Experimental Results

C.1 Additional Image Representation Results

To complement the dataset-level results reported in the main manuscript, we provide qualitative reconstruction comparisons on Kodak image 20 and an additional image from the DIV2K dataset. All methods are evaluated under the same image-representation protocol used in the main experiments.

Kodak image 20. Figure 12 presents reconstructions of Kodak image 20 at its native resolution. CoRe-INR achieves 46.80 dB, exceeding COSMO-RC by 1.70 dB. The enlarged wheel region further illustrates that CoRe-INR better preserves fine structures such as the tire tread, wheel spokes, and surrounding edges, whereas the lower-performing baselines exhibit increasing smoothing and loss of detail.

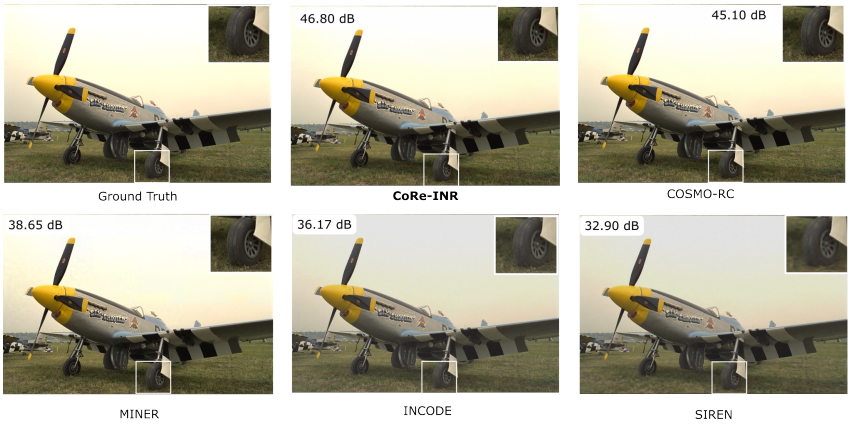


Figure 12: Qualitative image-representation results on Kodak image 20.

DIV2K image. We additionally evaluate image representation on a DIV2K image (Timofte et al., 2018). Following the evaluation protocol used in the previous COSMO-RC study, the image is downsampled by a factor of four from 1644×2040 to 411×510. As shown in Fig. 13, CoRe-INR achieves 47.75 dB, improving upon COSMO-RC by 4.39 dB. CoRe-INR more faithfully reconstructs fine fur texture, whiskers, and stripe boundaries, while several competing methods produce visibly smoother reconstructions. This result provides additional evidence that coordinate-rescaled branches improve the recovery of high-frequency image content.

C.2 Additional Super-Resolution Results

We provide additional quantitative and qualitative super-resolution results to complement the experiments reported in the main manuscript. Figure 14 presents the qualitative comparison for the

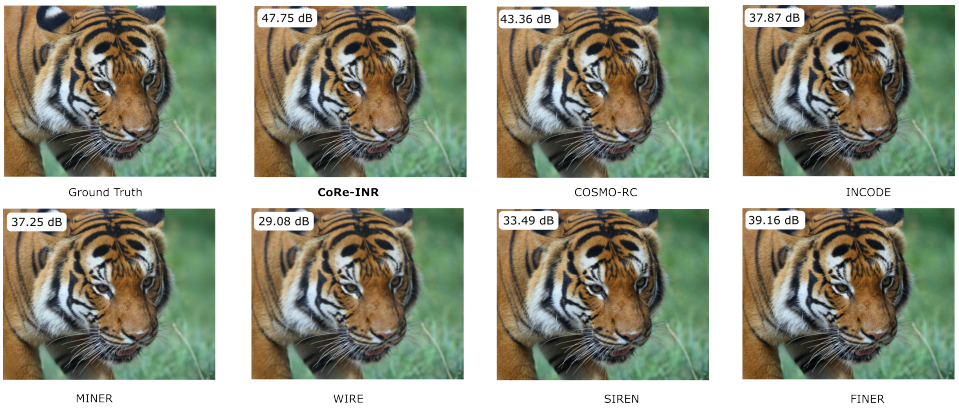


Figure 13: Qualitative image-representation results on a DIV2K image downsampled by a factor of four.

DIV2K $4\times$ setting evaluated in the main paper. CoRe-INR achieves 30.48 dB, marginally exceeding COSMO-RC at 30.42 dB, while preserving fine structures such as the butterfly-wing boundaries and surrounding flower details.



Figure 14: Qualitative $4\times$ super-resolution results on the DIV2K image evaluated in the main manuscript.

We further evaluate the methods on an image from the DIV4K-50 dataset (Zuo et al., 2025) at upsampling factors of $2\times$, $4\times$, and $6\times$. As shown in Table 5, CoRe-INR achieves the highest PSNR at all scales, outperforming the strongest competing method by 0.35 dB, 0.50 dB, and 0.78 dB, respectively.

Table 5: Super-resolution performance on the DIV4K-50 image.

METHOD	**2×**	**4×**	**6×**
	PSNR (dB) ↑	**PSNR (dB)** ↑	**PSNR (dB)** ↑
SIREN	27.55	27.22	26.73
INCODE	27.87	27.30	27.12
FINER	28.11	27.55	27.06
COSMO-RC	27.69	27.13	26.16
CoRe-INR	**28.46**	**28.05**	**27.90**



Figure 15: Qualitative $6\times$ super-resolution results on a DIV4K-50 image.

Figure 15 presents the corresponding qualitative comparison at $6\times$. CoRe-INR achieves 27.90 dB, exceeding the strongest baseline, INCODE, by 0.78 dB. It also better preserves facial contours, hair texture, the structure of the neon sign, and reflective details in the visor, whereas the competing methods exhibit greater smoothing and loss of detail.

C.3 Additional Image Denoising Results

To complement the denoising experiment reported in the main manuscript, we evaluate CoRe-INR on an additional image from the DIV2K dataset (Timofte et al., 2018). We use the same degradation process and training configuration as in the main experiment to ensure a consistent comparison across methods.

As shown in Fig. 16, CoRe-INR achieves 32.37 dB, outperforming the strongest baseline, COSMO-RC, by 0.51 dB. CoRe-INR better preserves the animal's facial markings, fur texture, harness boundaries, and background structure, whereas the competing methods exhibit greater smoothing and loss of fine detail. This additional example supports the denoising performance observed in the main manuscript under the evaluated noise configuration.

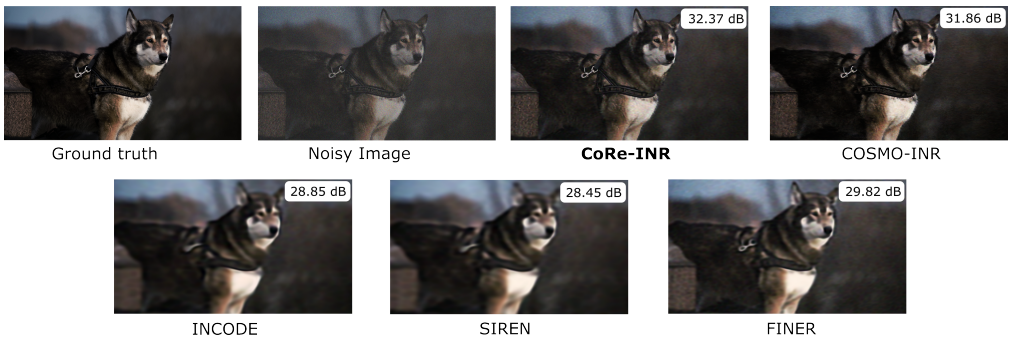


Figure 16: Additional qualitative image-denoising results on a DIV2K image.

D Efficiency and Computational Complexity

We evaluate all INR variants on Kodak image 20 (Franzen, n.d.) using the same NVIDIA T4 16 GB GPU and Adam optimizer. Table 6 reports parameter count, forward FLOPs, per-iteration training and inference time, inference throughput, and reconstruction PSNR.

The reported FLOPs include linear-layer operations but exclude activation-specific transcendental operations and memory overhead. CoRe-INR requires 534 K parameters and 69.5 forward GFLOPs because of its parallel coordinate-rescaled branches. Despite this higher linear-layer cost, it is substantially faster than COSMO-RC during both training (0.820 versus 3.500 s/it) and inference (0.291 versus 1.100 s/it), while achieving a throughput of 238.8 GFLOPs/s.

CoRe-INR attains the highest reconstruction quality at 46.8 dB, exceeding COSMO-RC by 1.7 dB and the strongest remaining baseline, FINER, by 10.4 dB. These results demonstrate that the additional branch-wise computation provides a favorable trade-off between computational efficiency and reconstruction quality.

Table 6: Computational efficiency and reconstruction quality of CoRe-INR and INR baselines on Kodak image 20.

METHOD	PARAMS (K)	FWD. GFLOPs	TRAIN (s/it)	INFER. (s/it)	THROUGHPUT (GFLOPs/s)	PSNR (dB)
SIREN	199	25.9	0.222	0.074	350.0	32.9
FINER	199	25.9	0.270	0.090	288.0	36.4
INCODE	437	38.7	0.435	0.145	267.0	36.2
WIRE	100	13.0	0.645	0.215	60.0	32.5
COSMO-RC	437	38.7	3.500	1.100	33.2	45.1
CoRe-INR	**534**	**69.5**	**0.820**	**0.291**	**238.8**	**46.8**