# C2F-INR: Leveraging Spectral Bias for Progressive Image Reconstruction

Sumit Kumar Dam[1] and Choong Seon Hong[2]
[1]*Department of Artificial Intelligence, Kyung Hee University, Yongin-si 17104, Republic of Korea*
[2]*Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, Republic of Korea*
E-mail: skd160205@khu.ac.kr, cshong@khu.ac.kr

*Abstract*—**Implicit Neural Representations (INRs) model images as coordinate-based continuous functions, offering a compact alternative to conventional pixel-grid representations. During training, these networks naturally progress from coarse-to-fine reconstructions, a behavior attributed to spectral bias that favors low-frequency components over high-frequency details. Rather than treating this tendency as a limitation, we harness it to design a structured learning framework. Inspired by progressive transmission in communication systems, where coarse visual content is delivered first and gradually refined with finer details, we propose a two-stage Coarse-to-Fine training strategy for INRs, referred to as C2F-INR, that explicitly governs this inherent progression. In the first stage, the network learns from a blurred target image to establish global structure, followed by a second stage that refines high-frequency details under the original target. Experiments on the KODAK dataset across two widely used INR backbones (SIREN and FINER) demonstrate consistent improvements in PSNR and SSIM, confirming the stability and efficiency of our proposed method. Thus, C2F-INR transforms the natural spectral bias into a controllable advantage, offering a new direction for progressive image reconstruction.**

*Index Terms*—**Implicit neural representation, progressive transmission, coarse-to-fine learning, image reconstruction**

## I. Introduction

Implicit Neural Representations (INRs) have recently emerged as a powerful paradigm for modeling complex signals in a continuous domain. Traditional representation methods rely on discretizing data into pixel grids, voxels, or point clouds, which limits their ability to capture the inherent continuity of real-world phenomena [1]–[9]. In contrast, INRs employ multilayer perceptrons (MLPs) to learn a mapping between spatial coordinates and corresponding signal values, effectively representing an image, a scene, or a shape as a continuous function rather than as a fixed-resolution array [10]–[17]. This coordinate-based formulation provides a compact and resolution-agnostic alternative to conventional formats, allowing reconstruction at arbitrary scales without additional memory overhead [18]–[20]. Besides, the continuous nature of INRs enables smoother interpolation and higher-quality reconstructions compared to discrete sampling approaches [2]. As a result, INRs have gained substantial attention across diverse domains, including image representation [21], [22], audio representation [15], [23], and 3D geometry modeling [24], [25]. Building upon this foundation, subsequent research has explored various network architectures and activation mechanisms to further enhance their representational capacity.

Recent developments in INRs largely focus on improving the fidelity of signal reconstruction through activation-based and architectural modifications. Among the early works, SIREN shows that using a sinusoidal activation enables smooth gradient propagation and precise modeling of complex structures [23]. Later, FINER extends this idea by allowing the activation frequency to vary adaptively during training, improving flexibility in representing different signal frequencies [10]. More recently, Fourier-ReLU (FR-INR) employs both sinusoidal and ReLU activations to maintain stability while capturing fine details [26]. Beyond these activation-focused studies, several works use positional encodings and multi-head structures to broaden the expressive range of INR frameworks [27], [28]. Despite their progress, these models still exhibit a natural tendency to reconstruct coarse components before finer ones, a behavior we reinterpret as a useful property rather than a drawback. In this work, we exploit this inherent progression to align with the principle of progressive transmission, where global information appears first and detailed content comes later.

Progressive transmission is a long-established concept in communication systems, where visual content is delivered in multiple stages, starting with a coarse preview that provides an early, recognizable image, followed by incremental refinements that improve perceptual quality. This step-wise delivery allows users to perceive meaningful content even before transmission is complete. Similarly, INRs exhibit a natural coarse-to-fine reconstruction behavior due to the spectral bias of neural networks. However, the early-stage outputs of standard INR models are often blurred and insufficiently detailed for practical use. To address this, we reinterpret this limitation as an opportunity for progressive reconstruction. Specifically, our two-stage Coarse-to-Fine INR (C2F-INR) framework biases early optimization toward global structural consistency by combining full-image supervision with coarse and edge guidance terms that gradually decay during training. This design produces perceptually coherent intermediate reconstructions at early epochs and transitions smoothly to fine-detail refinement, yielding stable and high-quality image reconstruction suitable

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00352423) and part by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing) and part by IITP(Institute of Information & Coummunications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT) (IITP-2025-RS-2023-00258649) *Dr. CS Hong is the corresponding author.

979-8-3315-7896-1/26/$31.00 ©2026 IEEE

for progressive-transmission applications. Overall, our contributions are summarized as follows:

- We propose a two-stage Coarse-to-Fine (C2F) training framework that explicitly structures the natural coarse-to-fine learning tendency of INRs into a guided optimization process. Rather than counteracting spectral bias, the framework turns it into an advantage through coarse and edge guidance terms that gradually decay over the course of training.
- The proposed C2F-INR produces perceptually coherent intermediate reconstructions by stabilizing global structure in the early epochs and gradually refining intricate details as training progresses. This property makes the model particularly suitable for progressive transmission, as it provides visually meaningful results even at intermediate epochs.
- Experimental results on the KODAK dataset [29] demonstrate consistent improvements in PSNR and SSIM over baseline INR models such as SIREN [23] and FINER [10], especially at early epochs (e.g., 100–300), confirming the efficiency, stability, and generality of the proposed framework.

The rest of the paper is organized as follows: Section II provides a review of the INR background, baseline models, and evaluation metrics. In Section III, we discuss our proposed methodology. Section IV presents the experimental results, demonstrating the effectiveness of our approach. Finally, Section V concludes the paper with a summary of our findings.

## II. Preliminary

### A. Background for Image INR

INRs provide a coordinate-based framework for representing images as continuous functions rather than discrete pixel grids. Instead of storing pixel values directly, INRs learn a mapping between pixel coordinates and their corresponding intensity or color values through a neural network. Formally, given a pixel coordinate $\mathbf{p} = (x, y)$ and its associated color vector $I(\mathbf{p}) \in [0, 1]^3$, an INR models this mapping as:

$$I(\mathbf{p}) = f_\theta(\mathbf{p}), \tag{1}$$

where $f_\theta$ denotes a neural network parameterized by $\theta$. The goal is to approximate the underlying continuous image signal by minimizing the reconstruction error between the predicted and target pixel values.

Typically, $f_\theta(\cdot)$ is implemented as a Multi-Layer Perceptron (MLP) that transforms input coordinates into pixel values through a series of linear and non-linear operations. For an $L$-layer MLP, the forward propagation of activations can be expressed as:

$$\mathbf{h}_{l+1} = \sigma(\mathbf{W}_l \mathbf{h}_l + \mathbf{b}_l), \tag{2}$$

where $\mathbf{W}_l$ and $\mathbf{b}_l$ represent the weights and biases of the $l$-th layer, and $\sigma(\cdot)$ is a non-linear activation function such as ReLU or Sine. The input to the network is the pixel coordinate $\mathbf{h}_0 = \mathbf{p}$, and the final output $\mathbf{h}_L$ corresponds to the predicted pixel value $\hat{I}(\mathbf{p})$. The optimization objective of image INR can be formulated as:

$$\theta^* = \arg\min_\theta \mathcal{L}(f_\theta(\mathbf{p}), I(\mathbf{p})), \tag{3}$$

where $\mathcal{L}$ represents a reconstruction loss, typically the mean squared error (MSE) between the predicted and ground-truth pixel intensities. The parameters $\theta$ are updated iteratively through gradient descent as:

$$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}(\theta), \tag{4}$$

where $\eta$ denotes the learning rate. Through this process, the network gradually learns to represent the image as a continuous signal that can be queried at arbitrary spatial coordinates, enabling smooth interpolation and high-fidelity reconstruction.

### B. Backbone Architectures

We employ SIREN [23] and FINER [10] as our backbone architectures to evaluate the effectiveness of the proposed framework. Both architectures are widely adopted in INR studies for their capability to model continuous signals and capture diverse frequency components.

**SIREN.** SIREN [23] models a continuous image function as:

$$f_\theta : (x, y) \to (r, g, b), \tag{5}$$

where $f_\theta$ denotes a multilayer perceptron (MLP) equipped with sinusoidal activation functions defined as:

$$\phi(z) = \sin(\omega_0 z), \tag{6}$$

with $\omega_0$ representing a frequency scaling factor that controls the periodicity of the sine function. The network learns to approximate the ground-truth image $I_{\text{gt}}$ by mapping sampled 2D coordinates to their corresponding RGB values. For each pixel $(i, j)$, the normalized spatial coordinate $(x_i, y_j)$ is used as input to the network. The network then predicts the corresponding pixel intensity as:

$$I_{\text{pred}}(i, j) = f_\theta(x_i, y_j). \tag{7}$$

This formulation enables smooth gradient propagation and facilitates accurate modeling of complex image structures.

**FINER.** FINER [10] extends the SIREN framework by introducing a frequency-adaptive activation function that evolves during training. Unlike SIREN, which employs a fixed frequency parameter, FINER dynamically adjusts the activation frequency, formulated as:

$$\phi(z) = \sin(\omega z), \tag{8}$$

where the frequency term $\omega$ varies throughout the optimization process. This adaptive mechanism allows the network to flexibly represent both low- and high-frequency signal components, thereby improving reconstruction fidelity and training stability across different frequency regions.

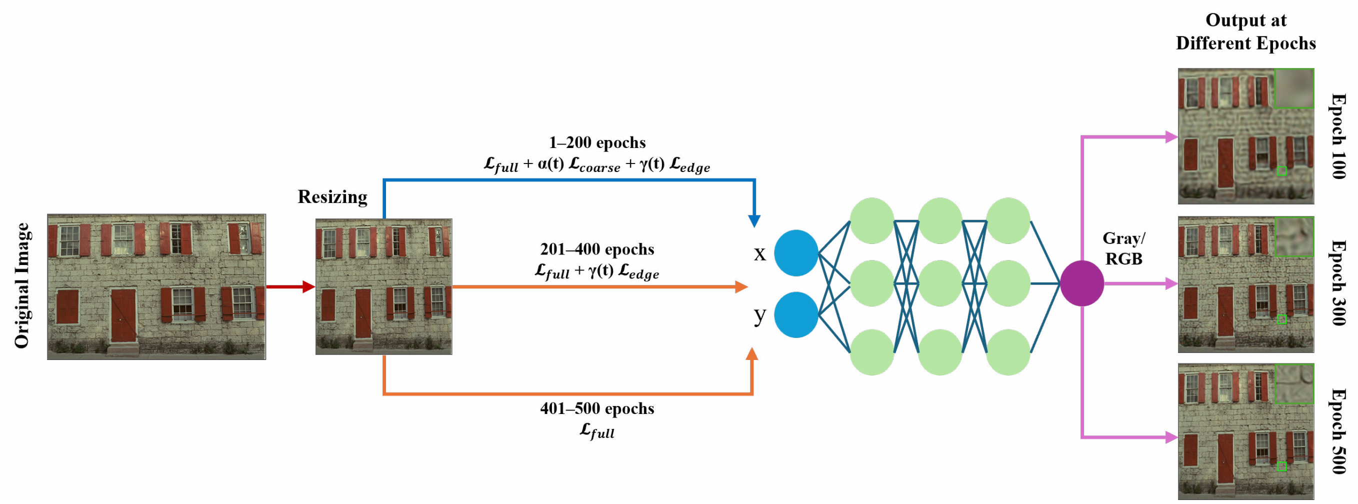

Fig. 1. Overview of the proposed two-stage Coarse-to-Fine INR (C2F-INR) framework. Stage 1 (1–200 epochs) jointly applies all three losses $\mathcal{L}_{\text{coarse}}$, $\mathcal{L}_{\text{edge}}$, and $\mathcal{L}_{\text{full}}$ for stable global and structural learning. Stage 2 (201–400 epochs) retains $\mathcal{L}_{\text{full}}$ and $\mathcal{L}_{\text{edge}}$, while the final fine-tuning phase (401–500 epochs) optimizes only $\mathcal{L}_{\text{full}}$ to maximize high-frequency detail and overall PSNR.

### *C. Evaluation Metrics*

To quantitatively assess the reconstruction quality, we employ two widely used metrics: peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [30]. PSNR measures pixel-level fidelity between the reconstructed and reference images, while SSIM evaluates their perceptual similarity in terms of luminance, contrast, and structure.

The PSNR is defined as:

$$\mathrm{PSNR}(I,\hat{I}) = 10\log_{10}\left(\frac{L^2}{\mathrm{MSE}(I,\hat{I})}\right), \tag{9}$$

where $L$ denotes the dynamic range of pixel values (*e.g.*, $L{=}1$ for normalized images or $L{=}255$ for 8-bit images), and $\mathrm{MSE}(I,\hat{I})$ is the mean squared error between the reference image $I$ and the reconstructed image $\hat{I}$, defined as:

$$\mathrm{MSE}(I,\hat{I}) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\big(I_{ij} - \hat{I}_{ij}\big)^2. \tag{10}$$

Here, $H$ and $W$ represent the image height and width, respectively. A higher PSNR corresponds to a lower pixel-wise error and better reconstruction fidelity.

To evaluate perceptual similarity, we further compute SSIM between local patches $\mathbf{u}$ and $\mathbf{v}$ as:

$$\mathrm{SSIM}(\mathbf{u},\mathbf{v}) = \frac{(2\mu_{\mathbf{u}}\mu_{\mathbf{v}} + C_1)(2\sigma_{\mathbf{uv}} + C_2)}{(\mu_{\mathbf{u}}^2 + \mu_{\mathbf{v}}^2 + C_1)(\sigma_{\mathbf{u}}^2 + \sigma_{\mathbf{v}}^2 + C_2)}, \tag{11}$$

where $\mu_{\mathbf{u}}$ and $\mu_{\mathbf{v}}$ denote local means, $\sigma_{\mathbf{u}}^2$ and $\sigma_{\mathbf{v}}^2$ are local variances, and $\sigma_{\mathbf{uv}}$ is their covariance. The constants $C_1{=}(K_1L)^2$ and $C_2{=}(K_2L)^2$ stabilize the division, with typical values $K_1{=}0.01$ and $K_2{=}0.03$. SSIM ranges from 0 to 1, where higher values indicate stronger structural and perceptual similarity to the reference image.

## III. Methodology

INRs often require extensive iterations before the reconstructed images reach visual fidelity. In general, INRs tend to capture only the dominant features in the early stages of training, which gradually become sharper as training progresses. To better structure the learning process, we draw inspiration from progressive transmission in communication systems, where coarse information is delivered first, and finer details are transmitted later. Following this intuition, we propose a two-stage Coarse-to-Fine (C2F) training framework that enables the model to first learn a stable coarse representation of the image and then recover fine details during the later stages of training. An overview of the proposed framework is shown in Fig. 1.

### *A. Constructing Coarse Targets and Edge Cues*

Given an image $I \in [0,1]^{H\times W\times C}$, we derive two auxiliary signals that guide the learning process without changing the final target: a *coarse target* that captures the overall appearance and an *edge cue* that highlights boundary transitions.

**Coarse target.** Given an image $I \in [0,1]^{H\times W\times C}$, we form a coarse target by applying a Gaussian blur to suppress fine details while preserving the global layout:

$$I_{\text{coarse}} = \mathcal{G}_\sigma * I, \qquad \sigma = 1.2,\ \text{radius} = 3, \tag{12}$$

where $\mathcal{G}_\sigma$ denotes a Gaussian kernel and $*$ is convolution.

**Edge cue.** In parallel, we extract an edge cue that helps the model preserve structural sharpness. Let $\nabla_S(\cdot)$ denote the Sobel gradient operator and $\|\cdot\|_2$ the per-pixel magnitude. For each channel of the image, the gradient response is computed, and the results are aggregated by averaging across all channels as follows:

$$E(I) = \mathrm{mean}_c\big(\|\nabla_{\mathrm{S}}(I_{:,:,c})\|_2\big), \tag{13}$$

where $\mathrm{mean}_c(\cdot)$ denotes averaging over the channel dimension. This produces a single-channel map emphasizing prominent boundaries and intensity transitions. The resulting map acts as a lightweight regularization term, encouraging the network to maintain edge consistency during learning without changing the reconstruction target.

TABLE I

Quantitative results on the *kodim14* image from the KODAK dataset at 100, 300, and 500 epochs. Best results are highlighted in bold. The C2F variants generally outperform their baseline models, showing higher PSNR and SSIM in most cases. Although C2F-FINER shows a slight dip in PSNR at 100 epochs, its SSIM still improves, and both PSNR and SSIM surpass the baseline at 300 and 500 epochs. Notably, the gains at early stages for both C2F-SIREN and C2F-FINER (100 and 300 epochs) indicate faster and more stable convergence, while the final results at 500 epochs confirm sustained reconstruction quality.

| Model | Epoch 100 | | Epoch 300 | | Epoch 500 | |
|---|---|---|---|---|---|---|
| | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ |
| SIREN | 21.80 | 0.4401 | 24.60 | 0.6119 | 26.41 | 0.7153 |
| C2F-SIREN | **22.11** | **0.4625** | **24.79** | **0.6560** | **27.25** | **0.7618** |
| FINER | **23.40** | 0.5478 | 26.67 | 0.7409 | 29.00 | 0.8263 |
| C2F-FINER | 23.39 | **0.5638** | **26.98** | **0.7682** | **29.88** | **0.8566** |

**Algorithm 1** Two-Stage Coarse-to-Fine Training Framework

```
Input: Image $I$; INR $f_\theta$; Gaussian ($\sigma = 1.2$, radius $= 3$);
    total epochs $T$; decay parameters $\alpha_0 = 0.5$, $T_1 = 200$,
    $\gamma_{\text{start}} = 0.10$, $T_{\text{psnr}} = 400$.
Output: Trained parameters $\theta$.

Compute: $I_{\text{coarse}} \leftarrow \mathcal{G}_\sigma * I$, $E(I) \leftarrow \text{mean}_c(\|\nabla_S(I_{:,:,c})\|_2)$
for $t = 0$ to $T-1$ do
    $\hat{I} \leftarrow f_\theta(\Omega)$
    Compute $\mathcal{L}_{\text{full}}$, $\mathcal{L}_{\text{coarse}}$, $\mathcal{L}_{\text{edge}}$ as defined in Section III.
    $\alpha(t) \leftarrow \alpha_0 \cdot \frac{1}{2}\big(1 + \cos\big(\pi \min(t, T_1)/T_1\big)\big)$
    if $t < T_{\text{psnr}}$ then
        $r \leftarrow \min(t/T_1, 1)$
        $\gamma(t) \leftarrow (1 - r) \cdot 0.05 + r \cdot \gamma_{\text{start}}$
    else
        $\gamma(t) \leftarrow 0$
    end if
    $\mathcal{L}_t \leftarrow \mathcal{L}_{\text{full}} + \alpha(t)\mathcal{L}_{\text{coarse}} + \gamma(t)\mathcal{L}_{\text{edge}}$
    $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_t$
end for
```

**Role in training.** The pair $\big(I_{\text{coarse}}, E(I)\big)$ provides complementary guidance throughout the learning process. The coarse target stabilizes early training by focusing on global structure, while the edge cue reinforces local boundaries and transitions. Their relative influence is adjusted dynamically during training, as detailed in the following subsection.

### *B. Two-Stage Training*

The training process is divided into two stages. In the **first stage**, the model is guided by both the coarse and edge components. This phase helps the network form a broad understanding of the global appearance while preserving key boundary information. By supervising the network with a coarse target, the optimization becomes less sensitive to pixel-wise noise and converges toward a smooth low-frequency structure.

Once the model has formed a stable representation, the **second stage** focuses on fine-detail reconstruction by discarding the coarse guidance and progressively reducing the edge weight, while continuing optimization with the original image target. This gradual transition helps the model retain structural sharpness from the first stage and refine high-frequency textures in the second stage.

Formally, the overall loss at epoch $t$ is expressed as:

$$\mathcal{L}_t = \beta\,\mathcal{L}_{\text{full}} + \alpha(t)\,\mathcal{L}_{\text{coarse}} + \gamma(t)\,\mathcal{L}_{\text{edge}}. \tag{14}$$

Here, $\mathcal{L}_{\text{full}}$ denotes the standard mean squared error between the prediction $\hat{I}$ and the original image $I$. It is defined as:

$$\mathcal{L}_{\text{full}} = \frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \|\hat{I}(i,j) - I(i,j)\|_2^2, \tag{15}$$

where $\Omega$ denotes the set of pixel coordinates in the image domain, and $\mathcal{L}_{\text{coarse}}$ and $\mathcal{L}_{\text{edge}}$ correspond to the auxiliary terms derived from the blurred target and the edge cue, respectively. The coefficients $\alpha(t)$ and $\gamma(t)$ control the weights of these terms over time. The coarse weight $\alpha(t)$ follows a cosine decay schedule that gradually reduces to zero after $T_1$ epochs and is defined as:

$$\alpha(t) = \alpha_0 \tfrac{1}{2}\big(1 + \cos\big(\pi \min(t, T_1)/T_1\big)\big)\,. \tag{16}$$

The edge weight $\gamma(t)$, on the other hand, starts with a small value to stabilize early optimization, gradually increases to enhance structural learning, and is finally reduced to zero in the fine-tuning phase for PSNR improvement. In practice, Stage 1 corresponds to $\alpha(t) > 0$ and $\gamma(t) > 0$, while Stage 2 begins with $\alpha(t) = 0$ and progressively reduces $\gamma(t)$ to zero, leading to the final reconstruction driven solely by $\mathcal{L}_{\text{full}}$. The complete procedure of the two-stage Coarse-to-Fine (C2F) training framework is summarized in Algorithm 1.

## IV. Experimental Results

To validate the effectiveness of the proposed C2F-INR framework, we conduct experiments on the KODAK dataset, where all images are resized to a resolution of $512 \times 512$. Each model is trained for 500 epochs using the Adam optimizer with a learning rate of $1 \times 10^{-4}$. The first 200 epochs emphasize coarse and edge-guided learning. The next 200 epochs focus on boundary refinement with edge consistency, and the final 100 epochs perform pure MSE-based fine-tuning for PSNR optimization. As mentioned earlier, SIREN and FINER are employed as the backbone architectures, and PSNR and SSIM are used as evaluation metrics. The proposed loss function integrates full MSE ($\mathcal{L}_{\text{full}}$), blurred-target MSE ($\mathcal{L}_{\text{coarse}}$), and Sobel-based edge consistency ($\mathcal{L}_{\text{edge}}$) with cosine-decayed

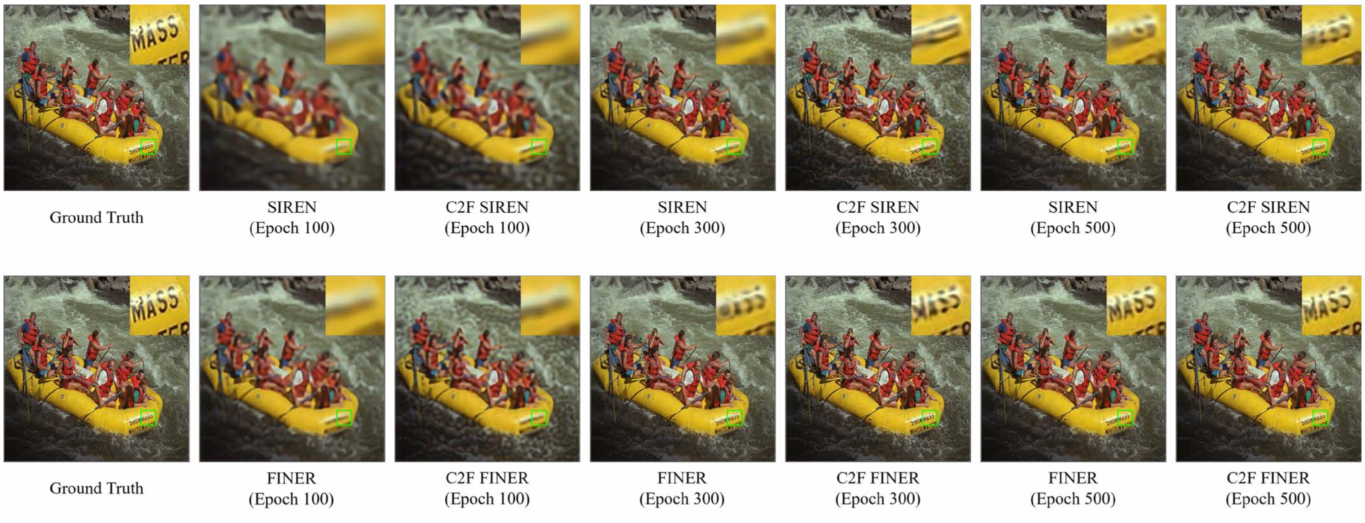

Fig. 2. Qualitative comparison between vanilla INR models and their C2F counterparts during a 500-epoch training. Results at 100 and 300 epochs demonstrate that the C2F variants achieve superior reconstruction quality with sharper edges and finer textures, indicating faster convergence in the early stages of training. By 500 epochs, the C2F variants maintain this visual superiority, highlighting the effectiveness of the proposed two-stage Coarse-to-Fine training framework.

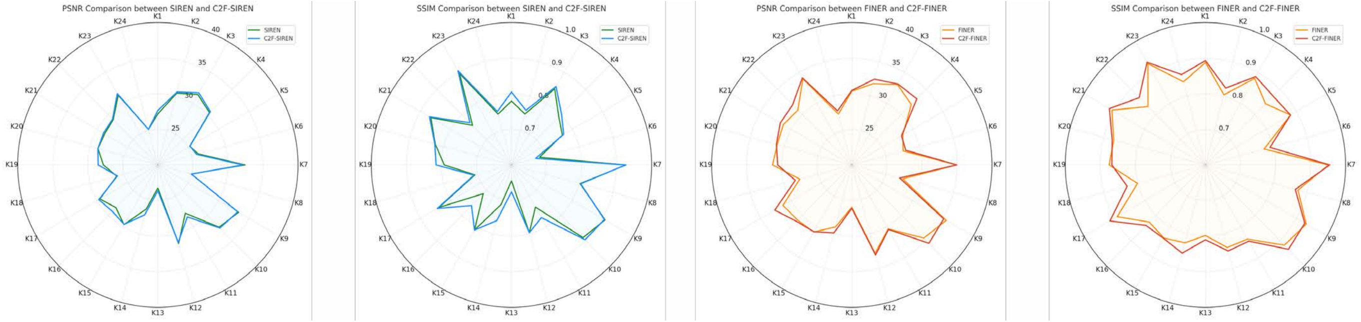

Fig. 3. Radar chart comparison of PSNR and SSIM across all 24 KODAK images for both SIREN and FINER, with and without the proposed C2F framework.

weighting, ensuring stable optimization and consistent convergence for both SIREN and FINER. All experiments are conducted on a single NVIDIA RTX A5000 GPU using the PyTorch 1.11.0 framework. To ensure a fair comparison, each INR model is configured with three hidden layers containing 256 neurons per layer. All other parameters follow the default configurations of their respective backbone implementations.

The quantitative results on the *kodim14* image from the KODAK dataset are presented in Table I. Across all epochs, the C2F variants show noticeable performance gains over their baseline counterparts. C2F-SIREN records consistent improvements of approximately +0.31 dB, +0.19 dB, and +0.84 dB in PSNR, and +0.0224, +0.0441, and +0.0465 in SSIM at 100, 300, and 500 epochs, respectively. C2F-FINER, on the other hand, shows a slight dip in PSNR at 100 epochs (–0.01 dB), although SSIM still improves by +0.016. Besides, at 300 and 500 epochs, it delivers clear gains of +0.31 dB and +0.88 dB in PSNR and +0.0273 and +0.0303 in SSIM, respectively. These results clearly indicate that our coarse-to-fine strategy enables faster convergence and better reconstruction at both early and later stages of learning. Notably, at 100 epochs, an early learning phase, the gains are more substantial, demonstrating that the coarse-to-fine learning scheme accelerates convergence by stabilizing gradient propagation and improving fidelity in both global and local structures. Even at 500 epochs, where models typically saturate, C2F-INR still produces higher PSNR and SSIM, validating the long-term consistency and efficiency of the coarse-to-fine learning strategy. Fig. 2 further confirms this trend, showing better reconstructions by the C2F variants compared to the baselines. Additionally, Fig. 3 provides a per-image radar chart comparison of PSNR and SSIM across all 24 KODAK images for both SIREN and FINER, with and without the proposed C2F framework. This visualization highlights the consistent improvements across most images.

TABLE II
AVERAGE PSNR AND SSIM ON THE KODAK DATASET. BEST RESULTS ARE HIGHLIGHTED IN BOLD.

| Model | PSNR ↑ | SSIM ↑ |
|---|---|---|
| SIREN | 28.63 | 0.7913 |
| C2F-SIREN | **28.82** | **0.8029** |
| FINER | 30.86 | 0.8608 |
| C2F-FINER | **31.17** | **0.8725** |

We also present the average quantitative results on the entire KODAK dataset in Table II. Consistent with the earlier findings, both C2F-SIREN and C2F-FINER outperform their baseline counterparts in terms of PSNR and SSIM. C2F-SIREN achieves a PSNR gain of +0.19 dB and an SSIM improvement of +0.0116 over the standard SIREN, while C2F-FINER shows an even larger improvement of +0.31 dB PSNR and +0.0117 SSIM compared to FINER. These results further confirm the generalizability of the proposed coarse-to-fine strategy, demonstrating that the structured training not only accelerates convergence at early epochs but also yields superior overall reconstruction quality.

## V. Conclusion

In this work, we propose C2F-INR, a two-stage Coarse-to-Fine training framework designed to leverage the natural spectral bias of implicit neural representations. Rather than treating the bias as a limitation, our approach instead turns it into an advantage by explicitly guiding the model from coarse structural learning to fine-detail refinement. Through a progressive training schedule and hybrid loss formulation, C2F-INR achieves faster convergence and produces superior reconstruction quality at both early and later stages of training. Experimental results on the KODAK dataset, using SIREN and FINER as backbones, demonstrate consistent improvements in PSNR and SSIM, validating the robustness and generality of the proposed C2F framework. In the future, we plan to extend this work to large-scale and dynamic signals, further exploring coarse-to-fine principles for real-time progressive reconstruction tasks.

## References

[1] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review*, vol. 40, p. 100379, 2021.

[2] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal representations on graphs: Tools and applications," *arXiv preprint arXiv:1512.05406*, 2015.

[3] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3d point clouds: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 12, pp. 4338–4364, 2020.

[4] Q. Dai, H. Chopp, E. Pouyet, O. Cossairt, M. Walton, and A. K. Katsaggelos, "Adaptive image sampling using deep learning and its application on x-ray fluorescence image reconstruction," *IEEE Transactions on Multimedia*, vol. 22, no. 10, pp. 2564–2578, 2019.

[5] Y. Sun, X. Tao, Y. Li, L. Dong, and J. Lu, "Hems: Hierarchical exemplar-based matching-synthesis for object-aware image reconstruction," *IEEE Transactions on Multimedia*, vol. 18, no. 2, pp. 171–181, 2015.

[6] T. Ogawa and M. Haseyama, "Missing image data reconstruction based on adaptive inverse projection via sparse representation," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 974–992, 2011.

[7] H. Wang and J. Zhang, "A survey of deep learning-based mesh processing," *Communications in Mathematics and Statistics*, vol. 10, no. 1, pp. 163–194, 2022.

[8] Y. Zang, B. Chen, Y. Xia, H. Guo, Y. Yang, W. Liu, C. Wang, and J. Li, "Lce-net: Contour extraction for large-scale 3d point clouds," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[9] P. An, Y. Duan, Y. Huang, J. Ma, Y. Chen, L. Wang, Y. Yang, and Q. Liu, "Sp-det: Leveraging saliency prediction for voxel-based 3d object detection in sparse point cloud," *IEEE Transactions on Multimedia*, 2023.

[10] Z. Liu, H. Zhu, Q. Zhang, J. Fu, W. Deng, Z. Ma, Y. Guo, and X. Cao, "Finer: Flexible spectral-bias tuning in implicit neural representation by variable-periodic activation functions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 2713–2722.

[11] S. Xie, H. Zhu, Z. Liu, Q. Zhang, Y. Zhou, X. Cao, and Z. Ma, "Diner: Disorder-invariant implicit neural representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6143–6152.

[12] V. Saragadam, J. Tan, G. Balakrishnan, R. G. Baraniuk, and A. Veeraraghavan, "Miner: Multiscale implicit neural representation," in *European Conference on Computer Vision*. Springer, 2022, pp. 318–333.

[13] L. Shen, J. Pauly, and L. Xing, "Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 770–782, 2022.

[14] E. Dupont, A. Goliński, M. Alizadeh, Y. W. Teh, and A. Doucet, "Coin: Compression with implicit neural representations," *arXiv preprint arXiv:2103.03123*, 2021.

[15] K. Su, M. Chen, and E. Shlizerman, "Inras: Implicit neural representation for audio scenes," *Advances in Neural Information Processing Systems*, vol. 35, pp. 8144–8158, 2022.

[16] C. Xu, J. Yan, Y. Yang, and C. Deng, "Implicit compositional generative network for length-variable co-speech gesture synthesis," *IEEE Transactions on Multimedia*, 2023.

[17] M. K. Suh, S. K. Dam, S. T. Kim, E.-N. Huh, and C. S. Hong, "Semantic-guided regularization to mitigate spectral bias in implicit neural representations," in *2025 25th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. IEEE, 2025, pp. 1–4.

[18] W. Fang, Y. Tang, H. Guo, M. Yuan, T. C. Mok, K. Yan, J. Yao, X. Chen, Z. Liu, L. Lu *et al.*, "Cycleinr: Cycle implicit neural representation for arbitrary-scale volumetric super-resolution of medical data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 11 631–11 641.

[19] M. Shao, C. Xia, D. Duan, and X. Wang, "Polarimetric inverse rendering for transparent shapes reconstruction," *IEEE Transactions on Multimedia*, 2024.

[20] D. Jayasundara, S. Rajagopalan, Y. Ranasinghe, T. D. Tran, and V. M. Patel, "Sinr: Sparsity driven compressed implicit neural representations," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 3061–3070.

[21] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8628–8638.

[22] S. Zheng, C. Zhang, D. Han, F. D. Puspitasari, X. Hao, Y. Yang, and H. T. Shen, "Exploring kernel transformations for implicit neural representations," *IEEE Transactions on Multimedia*, 2025.

[23] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.

[24] Z. Chen and H. Zhang, "Learning implicit fields for generative shape modeling," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5939–5948.

[25] J. Hu, K.-H. Hui, Z. Liu, R. Li, and C.-W. Fu, "Neural wavelet-domain diffusion for 3d shape generation, inversion, and manipulation," *ACM transactions on graphics*, vol. 43, no. 2, pp. 1–18, 2024.

[26] K. Shi, X. Zhou, and S. Gu, "Improved implicit neural representation with fourier reparameterized training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 25 985–25 994.

[27] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[28] A. Aftab, A. Morsali, and S. Ghaemmaghami, "Multi-head relu implicit neural representation networks," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 2510–2514.

[29] "Kodak lossless true color image suite," http://r0k.us/graphics/kodak/, accessed: 2025-10-30.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.