

CARA: Collision-Aware Resolution Adaptation for Multiresolution Hash Encoding Based Image Fitting

Linfeng Ye^{1,3}, Zhixiang Chi¹, Shayan Mohajer Hamidi², En-hui Yang³,
and Konstantinos N. Plataniotis¹

¹University of Toronto ²Stanford University ³University of Waterloo
linfeng.ye@mail.utoronto.ca; {zhxchi;kostas}@ece.utoronto.ca;
smohajer@stanford.edu; ehyang@uwaterloo.ca

Abstract. Multiresolution hash encodings have recently enabled fast and high-fidelity implicit neural representations by storing multi-scale features in fixed-size hash tables along a geometric resolution schedule. However, the standard design is data-agnostic: different resolution levels receive identical hash-table capacity despite large differences in image frequency content. As a result, some levels experience severe hash collisions while others underutilize parameters, leading to inefficient capacity allocation. To address this issue, we propose **C**ollision-**A**ware **R**esolution **A**daptation (CARA), a method that assigns per-level resolutions by balancing the effective information load across hash levels. This adaptive allocation reduces capacity bottlenecks and improves parameter efficiency. In addition, we introduce an invertible pixel-shuffle transform that reduces hash load factors by redistributing spatial information, thereby mitigating collision-induced information loss without enlarging the hash tables. To support evaluation on extremely high-resolution data, we also curate, to the best of our knowledge, the first uncompressed whole-slide image dataset for academic research. Experiments on Kodak images, gigapixel natural images, and raw whole-slide images demonstrate that CARA consistently improves the fidelity-parameter trade-off. Our method matches state-of-the-art performance while using only 27.76% of the parameters, and achieves up to 6.11dB PSNR improvement at comparable parameter counts. Code is provided in the supplementary.

Keywords: Implicit neural representations · Information allocation

1 Introduction

Implicit neural representations (INRs) [38, 42] have reshaped signal processing and computer vision by replacing discrete grid-based representations with continuous coordinate-based functional mappings. By parameterizing signals using deep neural networks [3, 4, 7], INRs provide a flexible and expressive framework for representing images. However, conventional multi-layer perceptrons (MLPs) often suffer from a spectral bias toward low-frequency components, which hinders the recovery of fine-grained, high-frequency details [27]. While architectural

techniques such as Fourier feature mappings and periodic activation functions partially alleviate this issue [24, 38, 42], focus has shifted toward hybrid grid-based representations, such as Instant-NGP [28, 41]. These methods accelerate optimization via multiresolution hash encoding, where spatial coordinates are mapped into a hierarchy of learnable hash tables and subsequently decoded by a lightweight MLP.

Despite their empirical success, existing multiresolution hash encoding-based methods suffer from two fundamental limitations as discussed in the sequel:

(i) *Data-agnostic resolution schedules*: They largely rely on data-agnostic encoding structures [28, 41, 43], which require manual hyperparameter tuning. In particular, they employ fixed, hand-crafted geometric resolution schedules that are shared across all images, regardless of spatial or temporal characteristics. However, images exhibit highly non-uniform information distributions across frequency bands (see Section 4.1). As a result, fixed resolution schedules may allocate excessive parameters to frequency regions with limited information content, while under-allocating capacity to information-dense regions that dominate perceptual error.

(ii) *Information degradation due to hash collisions*: Multiresolution hash encodings inherently suffer from hash collisions, where multiple grid vertices are mapped to the same table entry. The impact of such collisions varies significantly across resolution levels and depends on the underlying information density. Treating resolution assignment and collision behavior as independent design factors therefore leads to suboptimal parameter utilization.

We address both limitations with a unified framework. To address (i), we adapt per-level resolutions to balance the information load across frequency bands, preventing over-parameterization in sparse bands and under-capacity in information-dense bands, thereby significantly improving parameter efficiency. To address (ii), we apply an invertible pixel-shuffle transform that reduces hash collision by lowering the effective load factor before hashing.

We refer to the resulting framework as **collision-aware resolution adaptation** (CARA), which formulates resolution selection as an information-allocation problem. To evaluate the effectiveness of the proposed method, we conduct experiments on Kodak Lossless True Color Image Suite [12],¹ three gigapixel natural images (Pluto, Tokyo [26, 28], and Girl [28])² and four ultra-high resolution raw whole-slide images. Our contributions can be summarized as follows:

- We formulate resolution selection as an information-allocation problem to equalize the information encoded across layers, significantly improving parameter efficiency.
- We propose an effective pixel-shuffle approach to mitigate information loss caused by hash collisions.
- As elaborated in Sec. 5.3, we curate, to the best of our knowledge, the first *public uncompressed* whole-slide image (WSI) dataset for academic research, providing a challenging benchmark for high-resolution image fitting.

¹ We refer to this as the “Kodak dataset” hereafter.

² Pluto: Download link. Tokyo: Download link. Girl: Download link.

- Extensive experiments demonstrate that CARA outperforms state-of-the-art baselines by a large margin; for instance, achieving equivalent fidelity to previous state-of-the-art using only 27.76% of the parameters.

2 Related Works

Implicit neural representations (INRs) parameterize signals as continuous functions through deep neural networks [38, 42]. In its standard formulation, an INR takes a coordinate as input and produces the signal value at that location [40]. However, conventional MLPs exhibit a spectral bias toward learning low-frequency components first, which can hinder the efficient recovery of high-frequency details [38]. A large body of work mitigates this bias by proposing novel input encodings and activations, including Fourier feature mappings and periodic activations [24, 30, 32, 38], modifying the input coordinate representation [20, 42, 45], or increasing model capacity [14, 25]. These approaches expand the representable frequency content and improve optimization for fine-scale structures. Complementary to architectural changes, data-side transformations have been explored to improve INR fitting. Specifically, symmetric power transformations provide a reversible way to enforce range and approximate symmetry, demonstrating that input distributions materially affect INR optimization [57], and surprisingly, random pixel permutation [33] can accelerate the convergence.

Another line of work attempts to improve rendering quality by introducing learnable feature grids queried via interpolation, followed by a lightweight decoder [23, 28]. Overall, these approaches query points from a discrete grid and use the interpolated features as the input for a decoder. Various data structures have been proposed to improve efficiency, including 2D feature planes [31], sparse octree [51, 55], and hyperplanes [2, 13, 35, 37]. Such grid-based or hybrid representations substantially accelerate optimization and rendering. A particularly successful approach uses multiresolution hash-grid encoding [28], which stores features in hash tables across a geometric progression of resolutions. This significantly accelerates inference because the bulk of the signal information is stored in the grid, which can be accessed via fast $O(1)$ lookups.

Despite these advancements, most methods predominantly utilize data-agnostic data structures and necessitate manual parameter tuning to maximize reconstruction fidelity. Prior work typically relies on heuristic choices for the number of levels and geometric growth rates [10, 28].

This study investigates grid resolution selection and collision modeling for multiresolution hash encoding. Instead of applying a fixed geometric progression of resolutions, collision-induced information degradation is quantified at each resolution, and grid resolutions are selected to equalize the information encoded per level. This approach prevents excessive capacity allocation to frequency bands with limited information, thereby enhancing parameter efficiency and improving reconstruction fidelity.

3 Notation and Preliminaries

3.1 Notation

Scalars are denoted by non-bold letters (e.g., a, β), vectors by bold lowercase letters (e.g., $\mathbf{x}$), and tensors by bold uppercase letters (e.g., $\mathbf{Y}$). For a positive integer n , we write $[n] \triangleq \{1, \dots, n\}$. We use $\|\cdot\|$ for the Euclidean norm.

For a target image $\mathbf{Y}$ of spatial size $H \times W$, $\mathbf{Y}[i, j]$ denotes the entry at pixel index (i, j) . We associate pixel (i, j) with its normalized center coordinate

$$\mathbf{x}_{ij} \triangleq \left(\frac{j - \frac{1}{2}}{W}, \frac{i - \frac{1}{2}}{H} \right) \in [0, 1]^2, \quad (1)$$

so that all spatial frequencies are measured in cycles per pixel. Given an image tensor $\mathbf{Y}$ with intensity alphabet $\mathcal{V} \triangleq \{0, 1, \dots, 255\}$, we define its empirical entropy as

$$H(\mathbf{Y}) = - \sum_{v \in \mathcal{V}} \hat{p}(v) \log_2 \hat{p}(v), \quad (2)$$

where $\hat{p}(v)$ is the empirical probability mass function given by normalized histogram counts of v . For a multi-channel image, we compute entropy per channel and sum the results. For real-valued (e.g., filtered) signals, we apply a uniform quantizer before histogram estimation and compute $H(\cdot)$ on the quantized samples. Specifically, intensity-valued images use an 8-bit quantizer $Q_8(\cdot)$ with alphabet $\mathcal{V}$, whereas residual signals use a signed 9-bit quantizer $Q_\Delta(\cdot)$ with alphabet $\mathcal{V}_\Delta \triangleq \{-255, -254, \dots, 255\}$. Unless stated otherwise, $H(\cdot)$ denotes the empirical entropy computed over the corresponding fixed discrete alphabet.

3.2 Instant Neural Graphics Primitives

Implicit neural representations (INRs) model a discrete signal as samples of a continuous function $\mathbf{f}_\theta : \Omega \rightarrow \mathbb{R}^c$, where c denotes the number of output channels and $\mathbf{x} \in \Omega \subset \mathbb{R}^2$ represents spatial coordinates in the input domain. Given a set of training pairs $\{(\mathbf{x}_m, \mathbf{y}_m)\}_{m=1}^M$, the network parameters θ are optimized by minimizing a reconstruction loss,

$$\min_{\theta} \frac{1}{M} \sum_{m=1}^M \|\mathbf{f}_\theta(\mathbf{x}_m) - \mathbf{y}_m\|^2. \quad (3)$$

Coordinate-based MLPs can represent complex signals, but their optimization often suffers from spectral bias [27], exhibiting a preference for low-frequency components over high-frequency ones. This can slow convergence and reduce fidelity when the target contains rich high-frequency content [39, 42].

Instant neural graphics primitives (Instant-NGP) alleviates this limitation by combining a lightweight decoder with a trainable multiresolution hash encoding

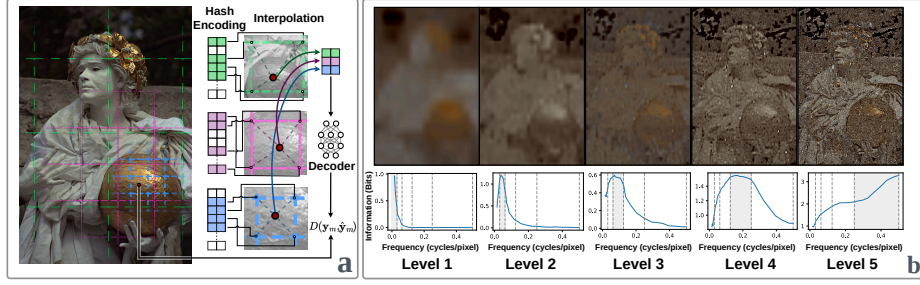


Fig. 1: Panel (a) illustrates that Instant-NGP factorizes an implicit neural representation into a multiresolution hash encoding and a lightweight MLP decoder. A query coordinate is interpolated within each grid level. The resulting per-level features are concatenated. The decoder is trained by minimizing the reconstruction distortion $D(\mathbf{y}_m, \hat{\mathbf{y}}_m)$. Panel (b) shows that after training an Instant-NGP model with a 5-level hash grid, we isolate the contribution of each level by retaining only the target level. All other levels’ hash-table entries are replaced with their mean feature value. The resulting reconstructions demonstrate clear spectral stratification. Lower-resolution levels recover smooth, low-frequency structures. Higher-resolution levels progressively contribute fine, high-frequency components. This behavior is quantified by the information density $\mathcal{I}_f$, defined in Equation (7), and plotted against frequency. Vertical dashed lines indicate the sampling frequency of each feature grid level. This confirms that different hash-grid levels encode information from different frequency bands.

that provides spatially localized, multi-scale features [28]. Specifically, Instant-NGP parameterizes an INR as

$$\mathbf{f}_\theta(\mathbf{x}) = \mathbf{g}_\theta(\phi(\mathbf{x})), \quad \phi(\mathbf{x}) = [\phi_1(\mathbf{x}); \dots; \phi_L(\mathbf{x})], \quad (4)$$

where $\mathbf{g}_\theta$ is a lightweight MLP and $\phi(\mathbf{x})$ concatenates the features from L resolution levels.

Assuming $\mathbf{x}$ is normalized to $\Omega = [0, 1]^2$, level ℓ defines a 2D grid with resolution N_ℓ per axis. Vertex embeddings are stored in a trainable table $\mathbf{E}_\ell \in \mathbb{R}^{T_\ell \times F}$, where T_ℓ is the table size and F is the feature dimension. A hash function $\mathbf{h}_\ell : \mathbb{Z}^2 \rightarrow [T_\ell]$ maps integer grid coordinates to table indices, enabling memory-efficient parameterization at high effective resolutions.

For a query $\mathbf{x}$, the level feature $\phi_\ell(\mathbf{x}) \in \mathbb{R}^F$ is obtained by bilinear interpolation of the embeddings at the four neighboring grid vertices:

$$\phi_\ell(\mathbf{x}) = \sum_{\delta \in \{0,1\}^2} w_{\ell,\delta}(\mathbf{x}) \mathbf{E}_\ell[\mathbf{h}_\ell(\lfloor N_\ell \mathbf{x} \rfloor + \delta)], \quad (5)$$

where $w_{\ell,\delta}(\mathbf{x})$ are the standard bilinear interpolation weights that sum to 1. Instant-NGP chooses the per-level resolutions using a data-agnostic geometric progression, $N_\ell = \lfloor N_{\min} b^{\ell-1} \rfloor$ for $\ell \in \{1, \dots, L\}$ and a predefined base $b > 1$. This yields grid spacing $\Delta_\ell \propto 1/N_\ell$, so higher levels (larger N_ℓ) can represent finer spatial variation. Concatenating features across levels provides

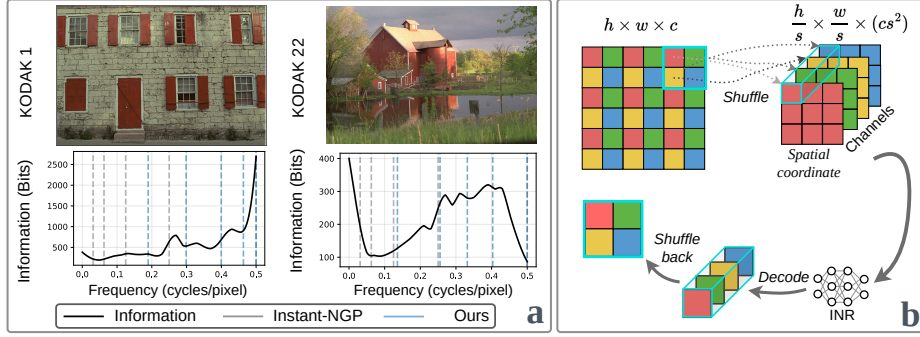


Fig. 2: Method overview. Panel (a): different images have different information density $\mathcal{I}(f)$, defined in Equation (7), for two Kodak images, showing that different images concentrate information at different frequency ranges; the vertical dashed lines indicate the frequency cutoffs implied by the fixed Instant-NGP schedule (gray) versus our collision-aware resolution schedule (blue), which aims to equalize the information being encoded by each level. Panel (b): pixel shuffling to mitigate collision-induced information loss. We partition the image into non-overlapping patches, then shuffle them into channels, thereby reducing the hash collision.

a multi-scale representation: low-resolution levels primarily capture coarse, low-frequency structure, while high-resolution levels encode finer, high-frequency details. As illustrated in Figure 1 panel (b), decoding with individual levels shows that higher-resolution grids recover high-frequency components, whereas lower-resolution grids represent the coarse structure of the signal.

4 Methodology

As discussed previously, the multiresolution hash encoding in Instant-NGP can be interpreted as a sampling-and-interpolation system [5]. At level ℓ , a grid with resolution N_ℓ per axis induces an effective sampling interval of $\Delta_\ell^{(x)} = w/N_\ell$ and $\Delta_\ell^{(y)} = h/N_\ell$ pixels along the horizontal and vertical directions, respectively. Under this interpretation, the corresponding Nyquist frequencies (in cycles per pixel) are

$$f_{\text{Nyq},\ell}^{(x)} = \frac{N_\ell}{2w}, \quad f_{\text{Nyq},\ell}^{(y)} = \frac{N_\ell}{2h}. \quad (6)$$

Frequencies beyond these limits cannot be represented without aliasing at level ℓ and must therefore be captured by finer-resolution levels or by the MLP decoder.

This observation suggests that each level ℓ primarily contributes to modeling a specific spatial frequency band determined by its resolution. Since frequencies are measured in cycles per pixel, a level with resolution N_ℓ per axis has an effective Nyquist limit $f_{\text{Nyq},\ell}^{(x)} = N_\ell/(2w)$ and $f_{\text{Nyq},\ell}^{(y)} = N_\ell/(2h)$. Hence, there exists an approximate correspondence between a cutoff frequency f and a grid resolution $N_f = 2wf$ (assuming square images for simplicity).

To quantify the information content associated with a frequency band centered at f , we measure the entropy [52, 54] increment between two adjacent spectral cutoffs.

Let $\mathbf{Y}_{\leq f}$ denote the low-pass filtered image retaining frequencies up to f (in cycles per pixel). We define the band information density as

$$\mathfrak{I}(f) = H(\mathbf{Y}_{\leq f+\Delta f} - \mathbf{Y}_{\leq f}) \cdot N_f^2, \quad (7)$$

where $H(\cdot)$ denotes the empirical Shannon entropy computed over pixel intensities, Δf is a small frequency increment (in cycles per pixel), and N_f is the grid resolution corresponding to cutoff frequency f . The multiplicative factor N_f^2 is the number of spatial samples at that resolution.

Remark 1 (Why entropy as an information proxy). Shannon’s source coding theorem [34, 49, 50, 53] implies that entropy quantifies the amount of information carried by a source. Accordingly, the quantity in Equation (7) can be interpreted as the new information introduced when increasing the effective cutoff from f to $f + \Delta f$ (measured in cycles per pixel). The increment $H(\mathbf{Y}_{\leq f+\Delta f} - \mathbf{Y}_{\leq f})$ captures the average additional uncertainty (bits per sample) contributed by this narrow frequency band, while N_f^2 denotes the number of spatial samples at the corresponding grid resolution. Their product therefore provides an estimate of the total information content of that refinement step. In this sense, $\mathfrak{I}(f)$ serves as a practical proxy for the information budget that a grid level should allocate in order to faithfully model structures up to cutoff frequency f . We ablate alternative information proxies in Section 6.2.

Equipped with the information density measure $\mathfrak{I}(f)$, we visualize its distribution for two representative images from the Kodak dataset [12] in Panel (a) of Figure 2. The two images exhibit markedly different spectral information distributions. Kodak 1 shows relatively limited low-frequency information, with a substantial portion of its entropy concentrated in higher-frequency bands. In contrast, Kodak 22 distributes its information more heavily across low- and mid-frequency regions, with comparatively reduced high-frequency content.

Despite such variability, Instant-NGP employs a data-agnostic geometric resolution schedule designed to uniformly cover a wide frequency range. In the next section, we introduce our method for selecting per-level resolutions adaptively based on the measured information distribution of the input.

4.1 Collision-Aware Resolution Adaptation

In this section, we present *collision-aware resolution adaptation (CARA)*, which assigns per-level grid resolutions according to the input’s measured information density. Our goal is to determine a resolution schedule $\{N_\ell\}_{\ell=1}^L$ such that each level is allocated a comparable effective information budget.

Let $\mathbf{Y}_{\leq f}$ denote the input signal low-pass filtered to cutoff frequency f (in cycles per pixel).³ Building upon the information density measure $\mathfrak{I}(f)$ defined

³ We study the impact of different low-pass filters in the supplementary material.

in Equation (7), we first aim to equalize the information allocated to each hash-grid level. To this end, we partition the frequency axis into L disjoint bands $\{\mathcal{F}_\ell\}_{\ell=1}^L$ such that each band carries the same total information:

$$\mathcal{I}_\ell \triangleq \sum_{f \in \mathcal{F}_\ell} \mathcal{J}(f) = \tau, \quad \forall \ell \in [L], \quad (8)$$

where τ is a shared target information budget. However, in hash encoding, the number of distinct trainable parameters at level ℓ is constrained by the hash table size T . Let $n_\ell = N_\ell^2$ denote the number of distinct grid vertices at that level. Since n_ℓ may exceed T , hashing induces collisions, where multiple grid vertices map to the same table entry. Assuming uniform hashing, the following proposition characterizes the expected reduction in usable capacity due to collisions.

Proposition 1 (Effective Capacity under Uniform Hash Collisions).

Suppose n_ℓ keys are inserted into a hash table of size T , with load factor $\alpha_\ell = n_\ell/T$. Under uniform hashing, the expected number of distinct entries per inserted key is

$$w_\ell^{\text{hash}} = \frac{1 - e^{-\alpha_\ell}}{\alpha_\ell}. \quad (9)$$

Accordingly, if B bits are stored across the n_ℓ keys, the expected retrievable information scales as

$$B_{\text{eff}} = B \cdot w_\ell^{\text{hash}}. \quad (10)$$

We defer the proof to the supplementary material due to the space limit. Following Proposition 1, we incorporate the collision factor into the per-level information by defining the effective information at level ℓ as

$$\mathcal{I}_{\text{eff},\ell} = \mathcal{I}_\ell \cdot w_\ell^{\text{hash}}. \quad (11)$$

After obtaining the collision-aware effective information, we determine the per-level resolutions so that different frequency bands carry comparable amounts of effective information.

Rather than discretizing frequency directly, we consider a monotone candidate set of grid resolutions $\{N^{(k)}\}_{k=1}^K$ with $N^{(1)} < \dots < N^{(K)}$, each implying a Nyquist cutoff in cycles per pixel. Let $Y^{(k)}$ denote the input low-pass filtered to the Nyquist limit corresponding to resolution $N^{(k)}$, and define $H_k \triangleq H(Y^{(k)})$, with H_0 a coarse reference. For each candidate k , we define the associated collision efficiency

$$w_k^{\text{hash}} \triangleq \frac{1 - e^{-\alpha_k}}{\alpha_k}, \quad \alpha_k \triangleq \frac{(N^{(k)})^2}{T}. \quad (12)$$

A resolution schedule is specified by indices

$$0 = k_0 < k_1 < \dots < k_L = K, \quad (13)$$

which partition the spectrum into L adjacent bands. The effective information carried by band ℓ , corresponding to the interval between cutoffs $k_{\ell-1}$ and k_ℓ , is defined as

$$\mathcal{I}_\ell \triangleq w_{k_\ell}^{\text{hash}} \sum_{f \in \mathcal{F}_\ell} \mathcal{J}(f), \quad (14)$$

which explicitly accounts for collision-induced capacity reduction at the selected resolution. We then select the cutoffs by solving the min-max optimization

$$\min_{0 < k_1 < \dots < k_{L-1} < K} \max_{\ell \in [L]} \mathcal{I}_\ell, \quad (15)$$

which seeks the tightest achievable uniform upper bound on per-band effective information.

Remark 2 (Why balance effective information and why min-max). Each hash-grid level has a bounded effective capacity: it uses a fixed-size table and its usable information is further reduced by collisions. Meanwhile, levels specialize to different frequency ranges (Figure 1 (b)), so reconstruction quality is often constrained by the most overloaded band: if one band contains much more information than its level can represent, the residual must be absorbed by other levels or the lightweight decoder, creating a capacity bottleneck, while over-allocating to a low-information band wastes parameters. We therefore balance per-band effective information as a load-balancing principle under a fixed per-level budget. Exact equality $\mathcal{I}_\ell = \tau$ is generally infeasible under discrete candidate cutoffs, so we minimize $\max_{\ell \in [L]} \mathcal{I}_\ell$ [1], *i.e.*, the tightest achievable uniform upper bound on per-band load, preventing any single band from becoming capacity-limiting.

The optimization problem in Equation (15) admits an exact solution via dynamic programming, as the objective decomposes over ordered band partitions with optimal substructure. We defer the detailed algorithm and pseudocode to the supplementary material. The resulting computational complexity is $O(LK^2)$, which is modest for typical candidate sizes K and negligible compared to the overall training cost of the INR, *e.g.*, adding only 0.7 seconds to pre-processing for the Pluto gigapixel image.

While CARA optimizes the allocation of grid resolution across frequency bands, the effective capacity of each level remains constrained by hash collisions. To further alleviate collision-induced information loss without increasing the hash table size, we introduce an efficient strategy dubbed pixel shuffling.

4.2 Reducing Hash Collisions via Pixel Shuffling

The performance of Instant-NGP improves as the hash table size T increases, since a larger table reduces the load factor and thereby improves the collision efficiency w^{hash} . However, enlarging T incurs a linear increase in both memory usage and the number of trainable parameters. To reduce collisions without increasing T , we apply the pixel shuffle operation [9, 36]. Given a target image

$\mathbf{Y} \in \mathbb{R}^{h \times w \times c}$ and an integer shuffle factor s , the pixel shuffle transform $\mathcal{S}_s$ rearranges $\mathbf{Y}$ into a lower-resolution, higher-channel representation

$$\tilde{\mathbf{Y}} = \mathcal{S}_s(\mathbf{Y}) \in \mathbb{R}^{\frac{h}{s} \times \frac{w}{s} \times (cs^2)}. \quad (16)$$

Each spatial coordinate in $\tilde{\mathbf{Y}}$ aggregates an $s \times s$ patch from the original image. Accordingly, the INR predicts a higher-dimensional vector

$$\hat{\mathbf{y}}_m = \mathbf{f}_\theta(\tilde{\mathbf{x}}_m) \in \mathbb{R}^{cs^2}, \quad (17)$$

and the full-resolution reconstruction is recovered via the inverse transform

$$\hat{\mathbf{Y}} = \mathcal{S}_s^{-1}(\hat{\tilde{\mathbf{Y}}}). \quad (18)$$

Importantly, pixel shuffling preserves all original information while reducing the spatial resolution of the coordinate domain by a factor of s along each axis. Consequently, at level ℓ , the number of distinct grid vertices decreases from N_ℓ^2 to $(N_\ell/s)^2$. Under a fixed hash table size T , the load factor becomes

$$\alpha'_\ell = \frac{(N_\ell/s)^2}{T} = \frac{\alpha_\ell}{s^2}. \quad (19)$$

By Proposition 1, reducing the load factor increases the collision efficiency factor w^{hash} , particularly when the original α_ℓ is large. Thus, pixel shuffling effectively improves usable capacity and rendering fidelity without increasing T . We defer the ablation study on s to the supplementary material.

5 Experiments

We evaluate the proposed method on three benchmarks: Kodak dataset, three gigapixel natural images, and ultra-high-resolution whole-slide images (WSIs). Unless otherwise specified, all models are trained with the Adam optimizer [19] with $(\beta_1, \beta_2, \epsilon) = (0.9, 0.99, 10^{-15})$, cosine annealing learning-rate decay, and an ℓ_2 reconstruction loss. Across all experiments, we employ the same multiresolution hash-grid architecture with feature dimension fixed to 2 at every level and the same number of levels for all methods and parameter budgets. To match



Fig. 3: Qualitative comparison on the Kodak dataset. By selecting per-level resolutions (rather than using a fixed, data-agnostic schedule), our method better preserves high-frequency structures and textures, yielding sharper reconstructions than I-NGP and MetricGrid. (Best viewed digitally with zoom).

Table 1: Image fitting results on the Kodak dataset. All methods are trained with the same number of iterations and a comparable number of parameters. Our method achieves the highest PSNR and SSIM while requiring fewer parameters compared to all other methods. **Bold** and underlined values denote the best and second-best results.

Method	WIRE [32]	SIREN [39]	SCONE [22]	I-NGP [28]	NFFB [48]	NeuRBF [6]	MetricGrid [46]	Ours
Rep.	Implicit	Implicit	Implicit	Hybrid	Hybrid	Hybrid	Hybrid	Hybrid
Params (K)	373	207	207	206	208	207	207	205
PSNR $\uparrow$	37.59	38.70	39.72	37.06	38.77	38.70	<u>39.73</u>	40.03
SSIM $\uparrow$	0.9596	0.9512	<u>0.9662</u>	0.9386	0.9537	0.9488	0.9568	0.9714

Table 2: Quantitative results of image fitting on the Kodak dataset. Comparing our method against the Gaussian primitive-based approach. **Bold** and underlined values denote the best and second-best results, respectively.

Method	I-NGP [28]	NeuRBF [6]	Gaussian Splatting [18]	GaussianImage [58]	MetricGrid [46]	Ours
Rep.	Hybrid	Hybrid	Explicit	Explicit	Hybrid	Hybrid
Params (K)	300	337	3540	560	335 533	341 471
PSNR $\uparrow$	43.88	43.78	43.69	44.08	44.51 45.77	<u>46.24</u> 48.00
SSIM $\uparrow$	0.9976	0.9964	0.9991	0.9985	0.9778 0.9897	0.9906 <u>0.9987</u>

different parameter budgets while keeping the architecture depth unchanged, we vary the hash-table size. For all experiments, we use anti-aliased resizing as the low-pass filter. For Kodak, we set $K = 30$ with uniform spacing in the frequency domain; for gigapixel images and WSIs, we set $K = 500$. Additional implementation details and wall-clock cost are provided in the supplementary material.

5.1 Kodak Dataset

We first evaluate our approach on the Kodak dataset [12], which consists of 24 natural images with resolution of 768×512 . We report PSNR and SSIM [47] to assess reconstruction fidelity. We compare against existing methods under comparable parameter budgets (around 207K), and train all models for the same number of iterations. Due to their relatively low resolution, we do not apply pixel shuffle to the Kodak dataset. The results are presented in Table 1. As seen, our method achieves the best image fidelity in terms of both PSNR and SSIM with fewer parameters.

We further compare our method against Gaussian primitive-based representations, including Gaussian Splatting (GS) [18] and GaussianImage [58]. To align the comparison across parameter regimes, we report CARA under two budgets in Table 2. At 471K parameters, CARA uses only 13% of the trainable parameters of GS (3540K) while achieving substantially higher PSNR. Moreover, CARA attains the best PSNR overall and improves upon the strongest hybrid baseline, MetricGrid [46], by 2.23 dB, highlighting the benefit of assigning resolutions based on the input image.

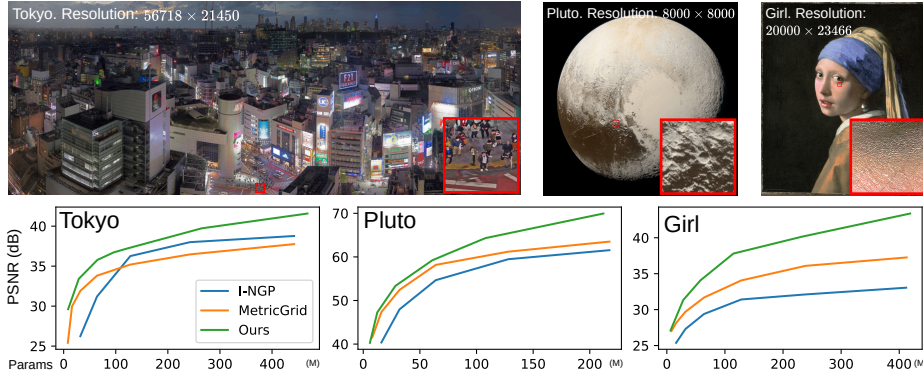


Fig. 4: Gigapixel Image Fitting Performance. **Top:** Gigapixel images (Tokyo, Pluto, and Girl) with magnified crops highlighting fine-grained details. (Best viewed digitally with zoom). **Bottom:** Quantitative comparison of reconstruction quality (PSNR) vs. the number of trainable parameters (M). Across all three images, our method consistently outperforms Instant-NGP (I-NGP) and MetricGrid, achieving a superior Pareto frontier and higher fidelity at every parameter scale.

5.2 Gigapixel Natural Images

To evaluate CARA’s performance at scale, we conducted experiments on three canonical gigapixel images: *Pluto*, *Tokyo*, and *Girl With a Pearl Earring* (Girl). These images pose a significant challenge for implicit neural representations, given their high resolution and the need to preserve fine-grained details across vast spatial domains. We compare the reconstructed image fidelity under multiple parameter budgets against Instant-NGP [28], NeuRBF [6], and MetricGrid [56]. Figure 4 reports PSNR versus the number of trainable parameters, where our method consistently achieves a better fidelity–parameter trade-off across all three images. Quantitatively, Table 3 shows that we match MetricGrid’s best reported fidelity (37.26 dB) using only 114.61M parameters, 27.76% of MetricGrid’s 412.8M parameter setting, demonstrating substantially improved parameter efficiency in the high-fidelity regime. Moreover, at a comparable large-model parameter budget, our method improves PSNR by 6.11 dB over the strongest baseline, demonstrating that collision-aware, data-adaptive resolution allocation yields consistent fidelity gains even in the high-capacity regime. Notably, in the low-parameter regime our method remains competitive, achieving higher image quality with 6.31M trainable parameters, which is 21.3% fewer than MetricGrid’s 8.02M setting. Finally, Figure 5 visualizes per-pixel ℓ_2 error maps on a gigapixel example; under matched training protocols, our approach yields visibly lower residual error than both baselines, corroborating the benefit of adaptive resolution allocation. Similar trends can be observed for Tokyo and Pluto; we defer the detailed analysis and visualization to the supplementary material due to space constraints.

Table 3: Numerical comparison on gigapixel girl image fitting. We report PSNR (dB), SSIM and the corresponding number of trainable parameters (in millions). Our method maintains competitive fidelity in both the low-parameter and high-parameter regimes.

Method	I-NGP [28]			NeuRBF [6]	MetricGrid [46]			Ours			
Params (M)	8.01	32.01	412.8	115.4	8.02	32.02	412.8	8.02	28.24	114.6	418.5
PSNR $\uparrow$	21.51	27.3	33.05	31.56	27.03	29.70	37.06	27.53	31.33	37.07	43.37
SSIM $\uparrow$	0.579	0.896	0.959	0.960	0.817	0.957	0.994	0.897	0.984	0.998	0.999

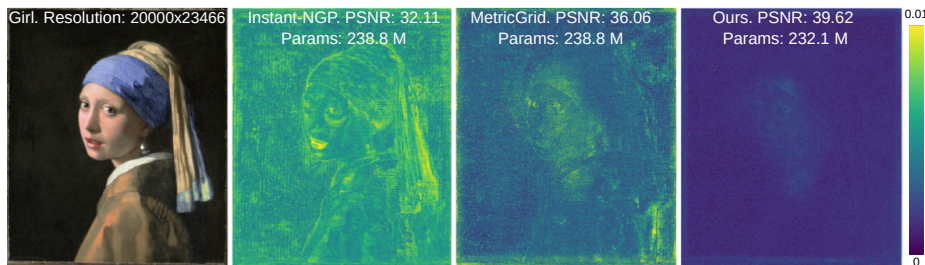


Fig. 5: Qualitative comparison on a gigapixel image. The leftmost presents the ground-truth image, while the three panels on the right visualize the per-pixel ℓ_2 error maps of I-NGP, MetricGrid, and the proposed method, respectively.

5.3 Whole-Slide Images

WSIs [16, 29] are ultra-high-resolution digital scans of histopathology slides, and constitute the primary data modality for computational pathology [15, 53]. Unlike natural images, WSIs contain micro-scale cellular morphology and macro-scale tissue architecture, requiring models to accurately represent structures across a wide spectrum of spatial frequencies. Their extreme spatial resolution and dense fine-grained texture pose significant challenges for conventional discrete representations.

In this section, we evaluate INRs to represent WSIs. Most publicly available WSI datasets are stored with aggressive JPEG compression [44], which introduces artifacts that can negatively affect downstream tasks [17]. To enable faithful evaluation, we collected four uncompressed *H&E* stained WSIs⁴ [11] at $40\times$ magnification (Figure 6 Panel (a)). Compared with compressed WSIs used in prior work (Figure 6 Panel (b)), our raw scans preserve fine structural detail. We benchmark CARA against four INR baselines and summarize the results in Table 4. CARA consistently outperforms the state-of-the-art methods across all parameter settings.

⁴ All patient-identifiable information has been rigorously removed to ensure privacy protection. The downsampled WSIs and dataset details are available in the supplementary material. The full-resolution images will be released upon acceptance.

Table 4: Numerical comparison on WSI image fitting. We report PSNR (dB) and the corresponding number of trainable parameters. Our method maintains competitive reconstruction fidelity in the low-parameter regime, while the high-capacity results are provided explicitly to facilitate direct numerical comparison in the large-model setting.

Method	I-NGP [28]			NeurRBF [6]	CINR [21]	MetricGrid [46]			Ours		
Params (M)	95.45	214.02	402.6	161.48	157.24	95.46	199.25	390.01	18.24	93.53	233.57
PSNR $\uparrow$	40.97	41.68	42.06	40.30	40.69	41.06	42.19	43.03	41.24	42.24	44.02

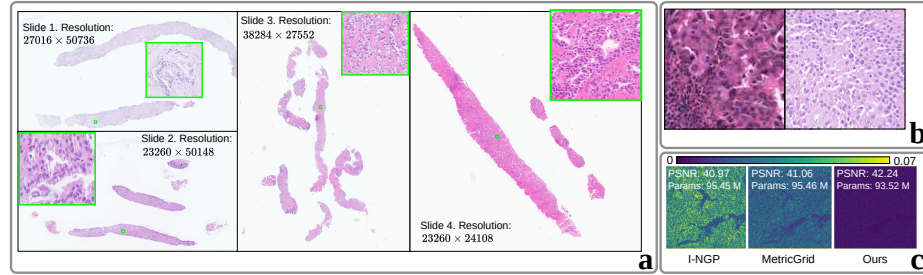


Fig. 6: WSIs used for evaluation. Panel (a) four ultra-high-resolution WSIs from our collection; green boxes indicate regions shown at higher magnification, highlighting both global tissue architecture and cell-level morphology. Panel (b) patch-level visual comparison between a WSI patch from prior literature (left) and a patch from our raw scans (right): aggressive JPEG compression in the literature data introduces visible blocking artifacts, whereas our raw WSIs preserve fine structural detail. (Best viewed digitally with zoom). Panel (c) per-pixel ℓ_2 reconstruction error maps on a representative patch; CARA produces consistently lower error, indicating improved image fidelity.

6 Ablation Study

To justify our design choices, we ablate each component in this section. Due to space constraints, additional results are provided in the supplementary material.

6.1 Effect of Collision Modeling

We ablate collision modeling by setting the collision factor to $w^{\text{hash}} \equiv 1$ while keeping the same entropy-based information estimator, hash-table size, number of levels, and training protocol. Table 5 shows that removing w^{hash} consistently

Table 5: Ablation of the collision factor w^{hash} on Kodak. Disabling collision modeling degrades PSNR/SSIM while typically requiring more parameters to reach comparable quality, highlighting the importance of explicitly accounting for hash collisions.

	W.O. Collision Factor			W. Collision Factor		
Params (K)	374	403	484	205	341	471
PSNR $\uparrow$	40.32	43.97	47.97	41.03	46.24	48.00
SSIM $\uparrow$	0.9674	0.9973	0.9979	0.9714	0.9906	0.9987

hurts reconstruction quality (PSNR/SSIM) and is less parameter-efficient, indicating that explicitly modeling collisions is a key component for allocating capacity effectively in hash-based encoders.

6.2 Alternative Information Proxies for Resolution Scheduling

Table 6: Effect of alternative information proxies for resolution scheduling on Kodak. We replace the entropy-based band information in Equation (7) with an ℓ_2 energy proxy (Ours + ℓ_2); variance proxy (Ours + var) and compare against the entropy proxy (Ours + entropy); entropy consistently yields higher PSNR/SSIM.

Method	I-NGP [28]		MetricGrid [46]		Ours+ ℓ_2		Ours+var		Ours+entropy	
Params (K)	206	300	207	335	214	334	211	331	205	332
PSNR $\uparrow$	37.06	43.88	39.73	44.51	38.52	43.60	38.74	44.94	41.03	46.63
SSIM $\uparrow$	0.9386	0.9964	0.9568	0.9978	0.9504	0.9913	0.9560	0.9893	0.9714	0.9982

As described in Section 4, our method uses empirical entropy to quantify the effective information in each frequency band. In this ablation, we replace the entropy term in Equation (7) with an ℓ_2 energy proxy or variance, and rerun the Kodak experiments under comparable parameter budgets. The results in Table 6 show a clear drop in both PSNR and SSIM when entropy is replaced by energy or variance. This behavior is expected: ℓ_2 energy characterizes average power but is largely insensitive to the underlying distribution of the band-residual signal, and variance is likewise a poor proxy for information. In contrast, entropy directly quantifies coding uncertainty, *i.e.*, the expected number of bits required to describe samples under an optimal code, thereby aligning with our goal of equalizing effective representational demand across levels.

7 Conclusion

In this paper, we presented collision-aware resolution adaptation (CARA), a data-adaptive framework for optimizing multiresolution hash encodings in implicit neural representations. By equalizing effective information across levels and explicitly accounting for collision-induced capacity reduction, CARA improved parameter utilization without increasing model size. We also introduced a pixel-shuffle strategy to further mitigate hash collisions in a parameter-efficient manner. Experiments showed consistent improvements in the fidelity-parameter trade-off across Kodak, gigapixel images, and whole-slide images, achieving comparable or better reconstruction quality with substantially fewer parameters. Limitations and future work are deferred to the supplementary material due to space constraints.

References

1. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge university press (2004)
2. Chan, E.R., Lin, C.Z., Chan, M.A., Nagano, K., Pan, B., De Mello, S., Gallo, O., Guibas, L.J., Tremblay, J., Khamis, S., et al.: Efficient geometry-aware 3d generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16123–16133 (2022)
3. Chen, A., Xu, Z., Wei, X., Tang, S., Su, H., Geiger, A.: Factor fields: A unified framework for neural fields and beyond. arXiv preprint arXiv:2302.01226 (2023)
4. Chen, H., He, B., Wang, H., Ren, Y., Lim, S.N., Shrivastava, A.: Nerv: Neural representations for videos. *Advances in Neural Information Processing Systems* **34**, 21557–21568 (2021)
5. Chen, Y., Wu, Q., Harandi, M., Cai, J.: How far can we compress instant-ngp-based nerf? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20321–20330 (2024)
6. Chen, Z., Li, Z., Song, L., Chen, L., Yu, J., Yuan, J., Xu, Y.: Neurbf: A neural fields representation with adaptive radial basis functions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4182–4194 (2023)
7. Chen, Z., Zhang, H.: Learning implicit fields for generative shape modeling. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5939–5948 (2019)
8. Chi, Z., Gu, L., Liu, H., Wang, Z., Wu, Y., Wang, Y., Plataniotis, K.: Learning to adapt frozen clip for few-shot test-time domain adaptation. In: International Conference on Learning Representations. vol. 2025, pp. 66359–66380 (2025)
9. Chi, Z., Wu, Y., Gu, L., Liu, H., Wang, Z., Zhang, Y., Wang, Y., Plataniotis, K.: Plug-in feedback self-adaptive attention in clip for training-free open-vocabulary segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 22815–22825 (2025)
10. Dai, T., Fan, J.: Characterizing and optimizing the spatial kernel of multi resolution hash encodings. In: The Fourteenth International Conference on Learning Representations (2026), <https://openreview.net/forum?id=q05hC1Pzkr>
11. Fischer, A.H., Jacobson, K.A., Rose, J., Zeller, R.: Hematoxylin and eosin staining of tissue and cell sections. *Cold spring harbor protocols* **2008**(5), pdb-prot4986 (2008)
12. Franzen, R.: Kodak lossless true color image suite (1999), <https://r0k.us/graphics/kodak/>
13. Fridovich-Keil, S., Meanti, G., Warburg, F.R., Recht, B., Kanazawa, A.: K-planes: Explicit radiance fields in space, time, and appearance. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12479–12488 (2023)
14. Hornik, K.: Approximation capabilities of multilayer feedforward networks. *Neural networks* **4**(2), 251–257 (1991)
15. Hosseini, M.S., Bejnordi, B.E., Trinh, V.Q.H., Chan, L., Hasan, D., Li, X., Yang, S., Kim, T., Zhang, H., Wu, T., et al.: Computational pathology: a survey review and the way forward. *Journal of Pathology Informatics* **15**, 100357 (2024)
16. Hosseini, M.S., Chan, L., Tse, G., Tang, M., Deng, J., Norouzi, S., Rowsell, C., Plataniotis, K.N., Damaskinos, S.: Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11747–11756 (2019)

17. Ignatov, A., Malivenko, G.: Nct-crc-he: Not all histopathological datasets are equally useful. In: European Conference on Computer Vision. pp. 300–317. Springer (2024)
18. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., et al.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
19. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
20. Landgraf, Z., Hornung, A.S., Cabral, R.S.: Pins: progressive implicit networks for multi-scale neural representations. *arXiv preprint arXiv:2202.04713* (2022)
21. Lee, D., Park, C., Lee, S., Lee, S., Kim, M.: Convolutional Implicit Neural Representation of pathology whole-slide images . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15007. Springer Nature Switzerland (October 2024)
22. Li, J.C.L., Liu, C., Huang, B., Wong, N.: Learning spatially collaged fourier bases for implicit neural representation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 13492–13499 (2024)
23. Liu, L., Gu, J., Zaw Lin, K., Chua, T.S., Theobalt, C.: Neural sparse voxel fields. *Advances in Neural Information Processing Systems* **33**, 15651–15663 (2020)
24. Liu, Z., Zhu, H., Zhang, Q., Fu, J., Deng, W., Ma, Z., Guo, Y., Cao, X.: 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. pp. 2713–2722 (2024)
25. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljacic, M., Hou, T.Y., Tegmark, M.: KAN: Kolmogorov-arnold networks. In: The Thirteenth International Conference on Learning Representations (2025), <https://openreview.net/forum?id=0zo7qJ5vZi>
26. Martel, J.N., Lindell, D.B., Lin, C.Z., Chan, E.R., Monteiro, M., Wetzstein, G.: Acorn: Adaptive coordinate networks for neural scene representation. *arXiv preprint arXiv:2105.02788* (2021)
27. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021)
28. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)* **41**(4), 1–15 (2022)
29. Omoush, S.A., Alzyoud, J.A., El-Omari, N.K.T., Alzyoud, A.J.: The role of whole slide imaging in ai-based digital pathology: current challenges and future directions—an updated literature review. *Journal of Molecular Pathology* **7**(1), 2 (2026)
30. Ramasinghe, S., Lucey, S.: Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In: European Conference on Computer Vision. pp. 142–158. Springer (2022)
31. Reiser, C., Szeliski, R., Verbin, D., Srinivasan, P., Mildenhall, B., Geiger, A., Barron, J., Hedman, P.: Merf: Memory-efficient radiance fields for real-time view synthesis in unbounded scenes. *ACM Transactions on Graphics (ToG)* **42**(4), 1–12 (2023)
32. Saragadam, V., LeJeune, D., Tan, J., Balakrishnan, G., Veeraraghavan, A., Baraniuk, R.G.: Wire: Wavelet implicit neural representations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18507–18516 (2023)

33. Seo, J., Lee, S., Kim, K.I., Lee, J.: In search of a data transformation that accelerates neural field training. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4830–4839 (2024). <https://doi.org/10.1109/CVPR52733.2024.00462>
34. Shannon, C.E.: A mathematical theory of communication. The Bell system technical journal **27**(3), 379–423 (1948)
35. Shi, K., Zhou, X., Gu, S.: Improved implicit neural representation with fourier reparameterized training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25985–25994 (2024)
36. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1874–1883 (2016). <https://doi.org/10.1109/CVPR.2016.207>
37. Shue, J.R., Chan, E.R., Po, R., Ankner, Z., Wu, J., Wetzstein, G.: 3d neural field generation using triplane diffusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20875–20886 (2023)
38. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
39. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 7462–7473. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper_files/paper/2020/file/53c04118df112c13a8c34b38343b9c10-Paper.pdf
40. Sun, J.M., Wu, T., Gao, L.: Recent advances in implicit representation-based 3d shape generation. Visual Intelligence **2**(1), 9 (2024)
41. Takikawa, T., Müller, T., Nimier-David, M., Evans, A., Fidler, S., Jacobson, A., Keller, A.: Compact neural graphics primitives with learned hash probing. In: SIGGRAPH Asia 2023 Conference Papers. pp. 1–10 (2023)
42. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. Advances in neural information processing systems **33**, 7537–7547 (2020)
43. Tang, J.H., Chen, W., Yang, J., Wang, B., Liu, S., Yang, B., Gao, L.: Octfield: Hierarchical implicit functions for 3d modeling. arXiv preprint arXiv:2111.01067 (2021)
44. Wallace, G.K.: The jpeg still picture compression standard. Communications of the ACM **34**(4), 30–44 (1991)
45. Wang, P.S., Liu, Y., Yang, Y.Q., Tong, X.: Spline positional encoding for learning 3d implicit signed distance fields. arXiv preprint arXiv:2106.01553 (2021)
46. Wang, S., Gao, Y., Li, S., Lv, C., Cai, X., Li, C., Yuan, H., Zhang, J.: Metricgrids: Arbitrary nonlinear approximation with elementary metric grids based implicit neural representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21381–21391 (2025)
47. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). <https://doi.org/10.1109/TIP.2003.819861>

48. Wu, Z., Jin, Y., Yi, K.M.: Neural fourier filter bank. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14153–14163 (2023)
49. Yang, E.H., Hamidi, S.M., Ye, L., Tan, R., Yang, B.: Conditional mutual information constrained deep learning: Framework and preliminary results. In: 2024 IEEE International Symposium on Information Theory (ISIT). pp. 569–574 (2024). <https://doi.org/10.1109/ISIT57864.2024.10619241>
50. Yang, E.H., Mohajer Hamidi, S., Ye, L., Tan, R., Yang, B.: Conditional mutual information constrained deep learning for classification. IEEE Transactions on Neural Networks and Learning Systems **36**(8), 15436–15448 (2025). <https://doi.org/10.1109/TNNLS.2025.3540014>
51. Yang, R.: Tinc: Tree-structured implicit neural compression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18517–18526 (2023)
52. Ye, L., Chi, Z.: Normalized conditional mutual information surrogate loss for deep learning classifiers. In: Workshop on Scientific Methods for Understanding Deep Learning
53. Ye, L., Hamidi, S.M., Chi, Z., Li, G., Pilanci, M., Ogawa, T., Haseyama, M., Plataniotis, K.N.: ASML: Attention-stabilized multiple instance learning for whole-slide imaging. In: The Fourteenth International Conference on Learning Representations (2026), <https://openreview.net/forum?id=CYmjrbQRyM>
54. Ye, L., Mohajer Hamidi, S., Tan, R., YANG, E.H.: Bayes conditional distribution estimation for knowledge distillation based on conditional mutual information. In: Kim, B., Yue, Y., Chaudhuri, S., Fragkiadaki, K., Khan, M., Sun, Y. (eds.) International Conference on Learning Representations. vol. 2024, pp. 26722–26754 (2024), https://proceedings.iclr.cc/paper_files/paper/2024/file/718a3c5cf135894db6e718725f52ef9a-Paper-Conference.pdf
55. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: Plenotrees for real-time rendering of neural radiance fields. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5752–5761 (2021)
56. Zhang, H.H., Zhang, T., Lin, B., Xue, Y., Zhu, Y., Liu, H., Gu, L., Ye, L., Wang, Z., Zuo, X., et al.: Widget2code: From visual widgets to ui code via multimodal llms. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20293–20302 (2026)
57. Zhang, W., Xie, S., Ren, C., Ge, S., Wang, M., Wang, Z.: Enhancing implicit neural representations via symmetric power transformation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 10157–10165 (2025)
58. Zhang, X., Ge, X., Xu, T., He, D., Wang, Y., Qin, H., Lu, G., Geng, J., Zhang, J.: Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. In: European Conference on Computer Vision. pp. 327–345. Springer (2024)