SymKL: A Unified Representation-Space Measure of Signal Degradation

Anonymous authors
Paper under double-blind review

Abstract

Quantifying degradation between a pristine signal and its degraded observation is fundamental to signal processing, yet existing full-reference measures are inherently modality-specific. We introduce **SymKL**, a unified full-reference degradation measure based on implicit neural representations (INRs). An INR is fitted once to the pristine reference and characterized by its diagonal Fisher information, which captures the local information geometry of the learned representation. Evaluating the Fisher information with the pristine reference and a degraded observation yields two parameter-sensitivity profiles whose symmetric KL divergence defines the degradation score. The proposed formulation requires only the reference and degraded observation, without knowledge of the corruption process or degradation severity, making it applicable across images, audio, and 3D data. We further derive an exact closed-form expression for the diagonal Fisher of SIREN, enabling efficient computation. Experiments on image, speech, and volumetric 3D benchmarks demonstrate that **SymKL** accurately tracks degradation severity, remains monotonic where conventional metrics saturate or fail, and provides a unified modality-agnostic measure of signal degradation.

1 Introduction

Signal degradation is an inevitable consequence of acquisition, transmission, compression, and reconstruction processes, making its accurate quantification a fundamental problem in signal processing. Reliable measures of reconstruction fidelity are indispensable for developing restoration algorithms, benchmarking competing methods, and evaluating reconstruction quality across a wide range of applications. Consequently, numerous degradation measures have been developed for different signal modalities. Images are commonly evaluated using pixel-fidelity measures such as PSNR (Huynh-Thu & Ghanbari, 2008), structural similarity measures including SSIM (Nilsson & Akenine-Möller, 2020) and MS-SSIM (Rouse & Hemami, 2008), and learned perceptual measures such as LPIPS (Zhang et al., 2018b) and DreamSim (Fu et al., 2023). Audio quality is typically assessed using signal-fidelity measures including SNR (Johnson, 2006) and SI-SDR (Le Roux et al., 2019), together with perceptual measures such as PESQ (Rix et al., 2001) and STOI (Zhang et al., 2018a), while 3D representations are commonly evaluated using geometric measures such as Chamfer Distance (Wu et al., 2021), Intersection-over-Union (IoU) (Zhou et al., 2019), and the F-score (Tatarchenko et al., 2019). Collectively, these

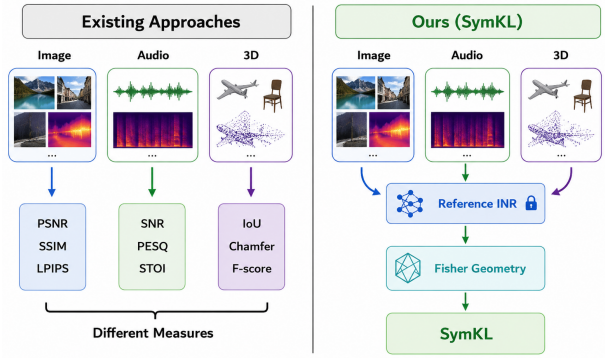

Figure 1: **Conceptual comparison between existing degradation measures and the proposed SymKL.** Existing approaches use modality-specific measures, whereas **SymKL** uses INR-based Fisher information geometry to provide a unified degradation measure across modalities.

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.

measures have become the standard tools for evaluating reconstruction quality within their respective domains.

Despite their success, these methods have largely evolved independently for different signal modalities and are built upon fundamentally different representations and similarity criteria. Image domain measures compare pixels, structural statistics, or learned visual features, audio measures evaluate waveform fidelity or perceptual intelligibility, while geometric measures assess spatial correspondence between shapes. Consequently, different modalities require fundamentally different degradation algorithms, each tailored to the characteristics of its underlying representation. Moreover, even within a single modality, the response of existing measures may become progressively less informative over parts of the degradation spectrum. For example, on the non-eccentricity distortion of KADID-10k (Lin et al., 2019), increasing the number of displaced image patches by a factor of five changes CLIP (Radford et al., 2021) similarity by less than 1% (from 0.998 to 0.990), while the visual distortion increases substantially. A similar phenomenon can be observed in 3D geometry, where the mesh F-score collapses to zero once the surface displacement exceeds its matching tolerance, despite geometric differences continuing to increase thereafter (See Supplementary). These observations motivate a fundamental question, *is there a unified representation in which signal degradation can be characterized consistently across different modalities while preserving sensitivity to distortion severity?*

Implicit neural representations (INRs) provide a promising foundation for addressing this question. Rather than representing signals as pixels, waveform samples, or geometric primitives, INRs learn continuous functions from discrete observations using a common coordinate-based neural network formulation. The same representation has been successfully applied to images, audio, videos, occupancy fields, signed-distance fields, neural radiance fields, and many other signal modalities. Unlike conventional degradation measures, which require different algorithms for different data types, INRs provide a unified representation capable of modeling a broad range of signals (Essakine et al., 2024; Jayasundara & Patel, 2026). This naturally leads to the central question of this work, *how is signal degradation encoded within this shared neural representation?*

Directly comparing INR parameters is uninformative because different parameters contribute unequally to representing the signal. Instead, **SymKL** compares their functional importance, quantified by the Fisher information (Ly et al., 2017), which measures the sensitivity of the reconstruction objective to parameter perturbations. A reference INR is fitted once to the pristine signal and then frozen. Evaluating its Fisher information using the pristine and degraded observations yields two parameter-sensitivity profiles whose discrepancy defines a modality-agnostic representation-space measure of degradation (see Figure 1).This naturally raises the question of why we characterize the representation using the Fisher information rather than gradients or the Hessian. The mean gradient vanishes near a fitted solution and therefore provides little information about the learned representation, while the Hessian, although capturing second-order curvature, can be computationally prohibitive for large INRs. The Fisher information retains the required second-order information while for SIREN networks, admits the exact closed-form expression.

Making the **SymKL** framework practical requires overcoming two computational challenges. Computing Fisher information through conventional automatic differentiation can be prohibitively expensive for large INRs. We therefore derive an exact closed-form expression for the diagonal Fisher information of SIREN networks (Sitzmann et al., 2020), reducing the computation to efficient matrix operations while preserving exactness. In addition, fitting a high-quality INR independently for every reference signal is computationally demanding. To address this limitation, we develop an amortized fitting framework based on a hypernetwork (Chauhan et al., 2024) with first-order MAML (Nichol et al., 2018) adaptation that rapidly produces accurate reference representations using only a small number of optimization steps. Finally, we introduce a physically grounded evaluation protocol that compares heterogeneous degradation measures according to their responsiveness to continuously varying physical distortion severity. Extensive experiments on natural images, audio, occupancy fields, and medical imaging demonstrate that the **SymKL** consistently tracks degradation severity, remains discriminative in regimes where conventional degradation measures become less informative, and substantially outperforms alternative weight-space characterizations.

Note that **SymKL** is not a perceptual quality metric. Unlike perceptual quality measures, our objective is not to approximate human judgments of visual or auditory quality. Instead, **SymKL** measures how strongly degradation alters the internal organization of a learned neural representation. Conse-

quently, it is applicable to signal modalities where perceptual judgments are ill-defined, unavailable, or not the primary objective, including volumetric representations, occupancy fields, medical imaging, and scientific measurements.

We make the following contributions. 1) We introduce the first modality-agnostic degradation measure based on the information geometry of implicit neural representations. 2) We show that degradation can be characterized through changes in parameter sensitivities rather than observation-space discrepancies. 3) We derive an exact closed-form Fisher computation for SIREN networks, enabling efficient evaluation. 4) We demonstrate a single degradation measure across images, audio, and 3D signals.

2 METHOD

2.1 PRELIMINARIES: SIGNALS AS COORDINATE NETWORKS

Let a signal be a mapping $x : \Omega \to \mathbb{R}^C$ on a coordinate domain $\Omega \subset [-1, 1]^D$, sampled at N coordinates $\{u_n\}_{n=1}^N$: an image (D=2, C=3), a waveform (D=1, C=1), or an occupancy/signed-distance volume (D=3, C=1). We represent x using a SIREN, an L-layer perceptron $f_\theta : \Omega \to \mathbb{R}^C$ with sinusoidal activations, where $h^{(0)} = u$, $z^{(\ell)} = W_\ell h^{(\ell-1)} + b_\ell$, and $h^{(\ell)} = \sin(\omega_\ell z^{(\ell)})$ for $\ell = 1, \ldots, L-1$, followed by the linear output layer $f_\theta(u) = W_L h^{(L-1)} + b_L$. The network parameters are $\theta = \{W_\ell, b_\ell\}_{\ell=1}^L \in \mathbb{R}^P$. The frequency scales ω_ℓ are modality-dependent (Sitzmann et al., 2020) ; importantly, the proposed degradation measure is identical across modalities. The INR is fitted by minimizing the mean per-coordinate reconstruction loss $\mathcal{L}(\theta; x) = \frac{1}{N}\sum_{n=1}^N \ell_n(\theta; x)$, where $\ell_n(\theta; x) = \frac{1}{2C}\|f_\theta(u_n) - x(u_n)\|_2^2$.

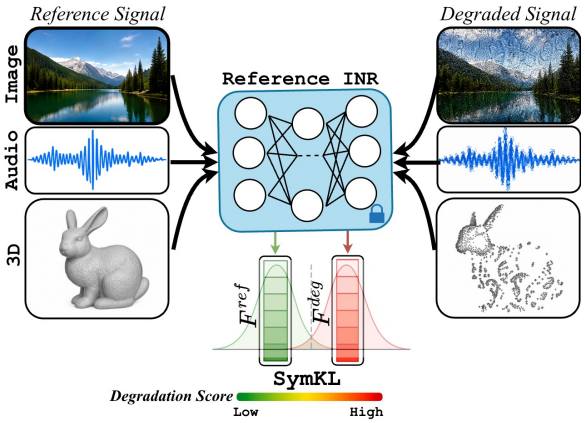

Figure 2: **Overview of `SymKL`.** Given a reference signal, **`SymKL`** first fits an INR to the reference and freezes its parameters. The frozen INR is then evaluated on both the reference and degraded signals to obtain their Fisher information representations. Their symmetric KL divergence yields a universal degradation score across diverse signal modalities.

2.2 AMORTIZED REFERENCE FITTING FOR IMAGES

Obtaining the reference INR through optimization can be computationally expensive (Sitzmann et al., 2020; Saragadam et al., 2023; Ramasinghe & Lucey, 2022). We therefore employ a hypernetwork Φ to predict the parameters of the target SIREN directly from the reference image, providing a high-quality initialization that substantially reduces the optimization required to obtain the reference representation. The proposed degradation measure is defined solely by the final reference representation and is therefore agnostic to how that representation is obtained. For modalities where no pretrained hypernetwork is available, such as audio and volumetric data, the reference INR is obtained through direct optimization.

The hypernetwork consists of three components: a DINOv2 ViT-B/14 encoder (Oquab et al., 2024), a transformer decoder, and lightweight prediction heads. Given a reference image, the encoder extracts dense patch features. The decoder maintains one learned query token per neuron of the target SIREN, which alternately attend to one another and cross-attend to the patch features, yielding a latent representation for each neuron. The prediction heads then map these latents to the corresponding weights and biases of the SIREN. We observed that directly predicting the network parameters is unstable, with unconstrained outputs the head activations grow without bound while the output scale shrinks to compensate, and training diverges (see Supplementary). Instead, the heads estimate bounded residuals with respect to a learnable base network, $\hat{W}_\ell = W_\ell^{\text{base}} + \beta_\ell^W \tanh(s\, o_\ell^W)$ and

$\hat{b}_\ell = b_\ell^{\text{base}} + \beta_\ell^b \tanh(s\, o_\ell^b)$, where o_ℓ^W, o_ℓ^b are the head outputs, $W_\ell^{\text{base}}, b_\ell^{\text{base}}$ are learnable base parameters initialized by the SIREN scheme, and s is a learned scale. The weight bounds β_ℓ^W are set to SIREN's own initialization bounds, and the bias bounds β_ℓ^b to one radian of phase, so predictions remain within a factor of two of a valid initialization. The heads are zero-initialized, so optimization begins exactly at the base network and learns bounded corrections to it rather than unconstrained parameters, substantially improving stability.

Training proceeds in two stages. We first pretrain the hypernetwork for image reconstruction on the combined MS-COCO (Awsaf, 2020) and LSDIR (Li et al., 2023) datasets, enabling it to predict high-quality initializations for the target SIREN. The network is then fine-tuned using a first-order MAML objective (Finn et al., 2017) so that the predicted parameters can be rapidly adapted to each reference image through a small number of reconstruction steps. During inference, the predicted parameters are refined. Additional architectural details, training schedules, optimization settings, and implementation details are provided in the supplementary material.

2.3 Fisher profiles of a frozen representation

Once the reference INR has been fitted, its parameters are frozen and are never updated again. Given this fixed representation, our goal is to characterize how differently it responds to the pristine signal and to its degraded counterpart, without learning a separate network for the degraded observation. Instead of comparing reconstructed signals or fitting a second INR, we measure how sensitive the parameters of the same reference representation are to each signal, quantified by the diagonal empirical Fisher information evaluated at the frozen parameters θ_{ref} (See Figure 2).

For any target signal y (either the pristine reference or a degraded observation defined on the same coordinates), let $g_n(y) = \nabla_\theta \ell_n(\theta; y)|_{\theta=\theta_{\text{ref}}}$ denote the reconstruction gradient at coordinate u_n. The gradient used for reconstruction optimization is the mean of these per-coordinate gradients, $\bar{g}(y) = \frac{1}{N}\sum_{n=1}^{N} g_n(y)$ (as in Preliminaries section). Instead of this first-order quantity, we consider their diagonal second moment, $F(y) = \frac{1}{N}\sum_{n=1}^{N} g_n(y) \odot g_n(y)$, corresponding to the diagonal empirical Fisher information. Intuitively, $F(y)$ measures the average squared sensitivity of each network parameter to the target signal. Near a fitted solution, the positive and negative contributions to the mean gradient largely cancel, causing $\bar{g}(y)$ to approach zero even though individual coordinates may still induce substantial parameter sensitivities. In contrast, the Fisher information accumulates squared gradients and therefore preserves the contribution of every coordinate. Empirically, we observe that $\|\sqrt{F}\|$ exceeds $\|\bar{g}\|$ by more than two orders of magnitude at θ_{ref}, illustrating the extent of this cancellation. Consequently, $F(y)$ provides a substantially richer per-parameter sensitivity profile for characterizing how the frozen representation explains the target signal. Following standard practice in continual learning (Kirkpatrick et al., 2017) and task embedding (Achille et al., 2019), we treat this diagonal profile as the precision of a Laplace-type Gaussian over the parameters, $p_y(\theta) = \mathcal{N}\big(\theta_{\text{ref}}, \operatorname{diag}(F(y))^{-1}\big)$. Crucially, all such distributions share the same mean θ_{ref}, since the reference representation is never updated.

2.4 The degradation measure

The previous section associates every target signal with a Fisher profile describing the parameter sensitivities of the frozen reference representation. Given the reference profile $F^{\text{ref}} = F(x_{\text{ref}})$ and the degraded profile $F^{\text{deg}} = F(x_{\text{deg}})$, our goal is to quantify how differently the two signals are represented. As each Fisher profile defines the precision of a Laplace Gaussian, this reduces to comparing two probability distributions over the same parameter space. A natural measure of the discrepancy between two probability distributions is the Kullback–Leibler (KL) divergence (Raiber & Kurland, 2017). For diagonal Gaussians with a shared mean, the KL divergence has the closed-form expression $\mathrm{KL}(F^{\text{ref}} \| F^{\text{deg}}) = \frac{1}{2P}\sum_{i=1}^{P}(r_i - 1 - \log r_i)$, where $r_i = F_i^{\text{deg}}/F_i^{\text{ref}}$. As the KL divergence is asymmetric, we define the final degradation measure as the symmetrized KL-divergence (**`SymKL`**), $\mathcal{D}(x_{\text{ref}}, x_{\text{deg}}) = \frac{1}{2}\mathrm{KL}(F^{\text{ref}} \| F^{\text{deg}}) + \frac{1}{2}\mathrm{KL}(F^{\text{deg}} \| F^{\text{ref}})$. The proposed measure satisfies $\mathcal{D} \geq 0$, with equality if and only if the two Fisher profiles are identical. Larger values indicate greater reorganization of the parameter-sensitivity structure required for the frozen representation to explain the degraded signal. Because the formulation depends only on the coordinate-based representation and the reconstruction objective, the same measure applies unchanged across different signal modalities.

2.5 EXACT CLOSED-FORM FISHER FOR SIREN

A direct evaluation of the empirical Fisher requires computing one parameter gradient for every coordinate, resulting in $N \times P$ gradient values for a signal with N coordinates and an INR with P parameters. For the signal resolutions considered in this work, for instance, corresponds to approximately 4×10^{10} values for a single image, making the naive computation impractical due to its memory, and time requirements. However, we observed that, the diagonal Fisher admits an exact closed-form computation for coordinate MLPs. Let $a_n^{(\ell-1)}$ denote the input activation to layer ℓ at coordinate n, and let $\delta_n^{(\ell)} = \partial \ell_n / \partial z_n^{(\ell)}$ denote the corresponding backpropagated error. The gradient of the loss with respect to the weight matrix is then the outer product $\partial \ell_n / \partial W_\ell = \delta_n^{(\ell)} (a_n^{(\ell-1)})^\top$. Consequently, its elementwise square is separable, allowing the summation over all coordinates to factorize exactly as $F_{W_\ell} = \frac{1}{N} \left(\Delta^{(\ell)} \odot \Delta^{(\ell)}\right)^\top \left(A^{(\ell-1)} \odot A^{(\ell-1)}\right)$, where the rows of $A^{(\ell-1)}$ and $\Delta^{(\ell)}$ contain the activations $a_n^{(\ell-1)}$ and backpropagated errors $\delta_n^{(\ell)}$, respectively, for all coordinates. The diagonal Fisher for the bias parameters follows analogously as $F_{b_\ell} = \frac{1}{N} \sum_n \delta_n^{(\ell)} \odot \delta_n^{(\ell)}$.The required backpropagated errors are obtained using the standard backward recursion, with the sinusoidal activation contributing only through its derivative $\omega_\ell \cos(\omega_\ell z_n^{(\ell)})$. Consequently, the complete Fisher is computed using one forward pass, one backward pass, and one matrix multiplication per layer, without ever materializing per-coordinate parameter gradients. The complete derivation is provided in Supplementary materials.

2.6 PHYSICALLY GROUNDED SENSITIVITY PROTOCOL

Beyond assigning a degradation score, a useful measure should respond meaningfully as the severity of the underlying physical distortion increases. The rate of this response reflects the measure's ability to distinguish progressively degraded signals. We therefore introduce a physically grounded sensitivity measure that quantifies the relative change in a degradation measure with respect to the relative change in the underlying physical distortion. As existing measures differ in polarity, scale, and interpretation, their responses are not directly comparable. We therefore map every metric m to a non-negative ratio-scale degradation value $\psi(m)$ with a true zero corresponding to identical signals. Similarity measures are transformed as $\psi(m) = 1 - m$, logarithmic fidelity measures such as PSNR, and SNR are converted to their corresponding linear error ratios using $\psi(m) = 10^{-m/10}$, while measures that already represent non-negative degradation or distance values are left unchanged. Given a monotonic sequence of distortion severity levels $d_1 < \cdots < d_K$ (e.g., increasing blur standard deviation, noise variance, erosion iterations, or reverberation time), and the corresponding transformed metric values $M_k = \psi(m_k)$, we define the step-wise sensitivity as $S_k = \left| \frac{(M_{k+1} - M_k)/M_k}{(d_{k+1} - d_k)/d_k} \right|$, which measures the percentage change in the degradation measure per percentage change in the underlying physical distortion. The overall sensitivity is obtained as the mean over all valid degradation steps, $S = \frac{1}{|\mathcal{V}|} \sum_{k \in \mathcal{V}} S_k$, where $\mathcal{V}$ denotes the set of valid consecutive steps. The same evaluation protocol is applied to every measure. Steps whose initial degradation value falls below a small noise threshold are excluded to avoid unstable ratios near zero, and distortion sequences that are non-monotonic in their physical parameters are evaluated as separate monotonic branches. This ensures that the reported sensitivity reflects the responsiveness of the degradation measure rather than artifacts of the evaluation protocol. Further details are provided in Supplementary materials.

3 EXPERIMENTS

We evaluate the proposed degradation measure on four representative signal modalities. RGB images, occupancy fields, audio, and medical imaging. These experiments demonstrate **SymKL**'s effectiveness in quantifying degradation across diverse domains while comparing against widely used modality-specific signal measures.

3.1 IMAGES

We evaluate how faithfully the proposed measure tracks degradation severity on the KADID-10k (Lin et al., 2019) dataset, which contains a diverse set of image degradations applied at multiple severity levels. For each reference image, an INR is fitted only to the pristine image using the

method discussed above, and subsequently frozen. The reference and its degraded versions are then evaluated using the same learned representation, ensuring that the observed differences arise from degradation rather than additional optimization. We compare against commonly used full-reference image-quality measures, including PSNR, SSIM, MS-SSIM, LPIPS, CLIP-based distance, and DreamSim. As some of these measures represent similarity while others represent distance, for a fair evaluation their outputs are converted to a common degradation-oriented form before computing sensitivity as explained in sensitivity protocol section. Beyond commonly used image-domain measures, we investigate several weight-space measures that can be computed directly from the frozen reference INR without re-optimizing it on the degraded image. This controlled setting isolates the choice of weight-space representation from the effects of additional optimization. Let F^{ref} and F^{deg} denote the diagonal Fisher profiles induced by the reference and degraded images, and let $\bar{g}^{\mathrm{ref}}$ and $\bar{g}^{\mathrm{deg}}$ denote their corresponding mean-gradient vectors. We consider five alternatives. **Fisher cosine (F-Cos)** computes the cosine distance between F^{ref} and F^{deg}, capturing changes in the relative distribution of Fisher values while remaining invariant to uniform scaling of their overall magnitudes. **Fisher relative ℓ_2 (F-ℓ_2)** computes the relative Euclidean difference, $\|F^{\mathrm{deg}} - F^{\mathrm{ref}}\|_2 / \|F^{\mathrm{ref}}\|_2$, making it sensitive to changes in both the relative distribution and overall magnitude of the Fisher values. **Gradient cosine (G-Cos)** applies cosine distance to the mean-gradient vectors, measuring differences in the directions of the first-order responses. **Gradient relative ℓ_2 (G-ℓ_2)** computes the relative Euclidean difference between $\bar{g}^{\mathrm{ref}}$ and $\bar{g}^{\mathrm{deg}}$, quantifying changes in the magnitude of the first-order response. Finally, **Adam-step cosine (Adam-Cos)** computes the cosine distance between two first-step Adam update vectors, one induced by the reference image and the other by the degraded image, both evaluated from the same frozen reference parameters with the optimizer state initialized to zero.

Table 1: Mean relative sensitivity on KADID-10k distortions image-domain and weight-space baselines. Best results are shown in bold and second-best results are underlined. Higher is better.

	Image-domain baselines						Weight-space baselines					
Distortion	**PSNR**	**SSIM**	**MS-SSIM**	**LPIPS**	**CLIP**	**DSim**	**F-Cos**	**F-ℓ_2**	**G-Cos**	**G-ℓ_2**	**Adam-Cos**	**SymKL**
Motion blur	1.678	1.659	2.678	1.187	2.399	2.860	0.391	2.749	0.345	1.137	0.132	**3.479**
JPEG2000	1.418	0.843	1.154	1.363	1.506	1.558	1.963	1.962	0.649	0.790	0.179	**2.921**
White noise	0.963	0.364	0.634	0.552	0.339	0.557	0.504	0.974	0.396	0.512	0.120	**1.169**
Impulse noise	1.004	0.799	0.827	0.515	0.334	0.702	0.461	1.004	0.452	0.580	0.118	**1.345**
Mean shift	3.059	2.382	3.222	3.172	2.787	3.338	0.402	2.990	0.052	1.068	0.034	**3.462**
Quantization	2.340	1.244	1.753	1.721	1.309	1.610	1.127	2.236	0.112	0.901	0.069	**2.862**
Contrast	3.468	3.046	3.815	3.227	3.318	3.663	0.413	3.804	0.076	1.195	0.048	**4.359**

Table 1 compares the relative sensitivity of the proposed measure against both image-domain and weight-space baselines on representative KADID-10k distortion types. As can be seen across the distortions, **SymKL** consistently exhibits the highest relative sensitivity, indicating a stronger and more consistent response to degradation severity than existing alternatives. These results suggest that modeling degradation through Fisher-profile distributions captures distortion-induced changes more effectively than conventional image-domain metrics or other weight-space statistics extracted from the same frozen INR.

Table 2: Responses of different image-quality metrics to progressively increasing color-block degradation (CB) and decreasing quantization levels (Q), both corresponding to increasing degradation severity. Underlined values indicate violations of the expected monotonic response.

	Color Blocks (CB)				Quantization (Q)			
Metric	**CB2**	**CB6**	**CB8**	**CB10**	**Q20**	**Q16**	**Q10**	**Q7**
PSNR	25.865	23.131	21.295	21.711	31.529	29.515	25.421	23.884
SSIM	0.989	0.970	0.964	0.958	0.956	0.932	0.856	0.864
MS-SSIM	0.981	0.920	0.874	0.917	0.990	0.983	0.957	0.959
LPIPS	0.045	0.135	0.196	0.166	0.017	0.029	0.078	0.060
CLIP	1.000	0.820	0.835	0.800	0.976	0.982	0.974	0.973
DreamSim	0.129	0.243	0.331	0.319	0.007	0.011	0.025	0.020
SymKL	**0.948**	**1.957**	**3.066**	**3.126**	**0.539**	**0.911**	**2.512**	**4.227**

To further examine whether quality measures respond consistently to increasing degradation severity, we analyze two progressive distortion sequences from the KADID dataset corresponding to color-block corruption and intensity quantization (Figure 3). The associated metric responses are summarized in Table 2. As each successive image represents a strictly stronger degradation than its predecessor, an ideal degradation measure should exhibit a monotonic response. However, several widely used image-quality metrics violate this expected behavior. For example, PSNR, SSIM,

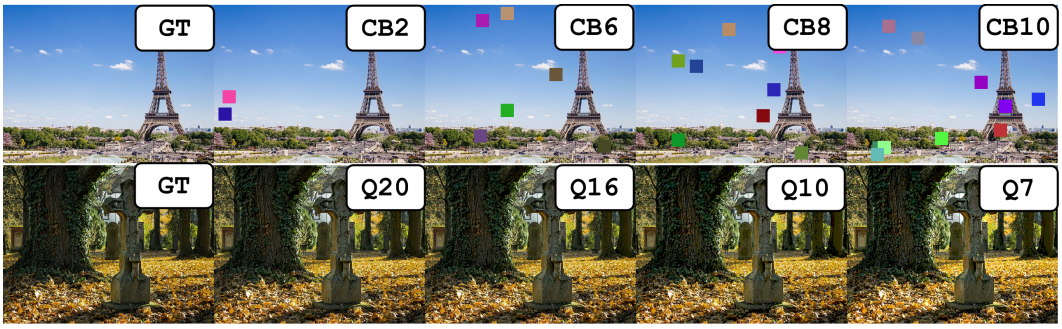

Figure 3: Responses of different image-quality metrics to progressively increasing degradation severity. **Top:** increasing color blocks (2–10). **Bottom:** decreasing quantization levels (Q20–Q7). Existing metrics exhibit inconsistent or non-monotonic responses, whereas **SymKL** increases smoothly and monotonically with degradation severity.

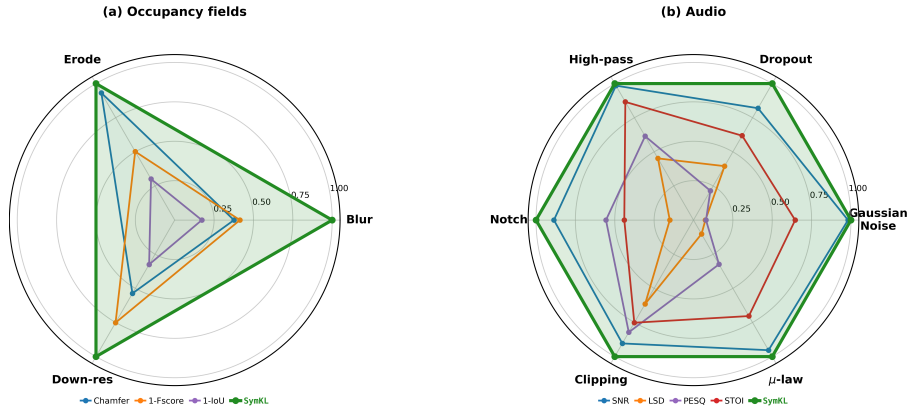

Figure 4: **Normalized sensitivity profiles across degradation types.** Each axis is normalized by the highest sensitivity attained for the corresponding degradation. Across both occupancy-field (left) and audio (right) degradations, **SymKL** consistently exhibits the strongest relative sensitivity, demonstrating a more responsive characterization of degradation than conventional modality-specific measures.

MS-SSIM, LPIPS, CLIP, and DreamSim all exhibit non-monotonic responses for at least one of the two degradation sequences, as indicated by the underlined entries in Table 2. In contrast, the proposed Fisher-based measure increases monotonically across both distortion sequences, producing a consistent response to progressively stronger degradations. These results suggest that the proposed representation provides a more stable and predictable characterization of degradation severity than both conventional image-domain measures and alternative learned perceptual metrics. Additionally, on a single NVIDIA RTX A5000 GPU, **SymKL** exhibits comparable evaluation time to DreamSim, requiring 52.36 ± 0.05 ms and 45.17 ± 0.74 ms per evaluation, respectively across three runs.

3.2 Occupancy Fields

To evaluate whether the proposed degradation measure generalizes beyond natural images, we next consider occupancy fields representing 3D objects. We first generate the reference occupancy fields by following the construction procedure described in WIRE (Saragadam et al., 2023). We then progressively degrade these reference occupancy fields using three commonly encountered degradation types, Gaussian blur, morphological erosion, and resolution reduction. Blur is applied using Gaussian kernels with increasing standard deviations ($\sigma = \{1.0, 1.5, 2.0, 3.0, 4.0\}$), erosion using 1–5 morphological iterations, and resolution reduction using downsampling factors of $\{1.5\times, 2\times, 3\times, 4\times, 6\times\}$. For each degradation sequence, we evaluate the response of the pro-

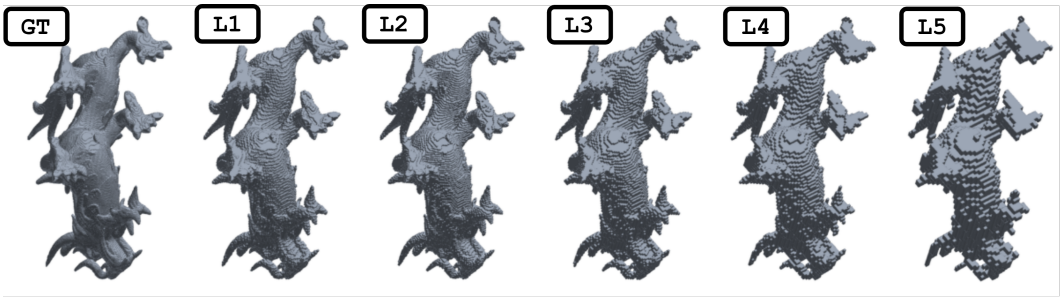

Figure 5: Progressive downsampling of a representative occupancy field. From `GT` to `L5` (DS6), increasing downsampling progressively removes fine geometric details while preserving the overall structure.

posed measure and compare it against the geometric metrics commonly used for occupancy-field evaluation, namely Chamfer distance, F-score, and volumetric IoU. Left of Figure 4 illustrates the max-normalized mean relative sensitivity over all degradation levels for the *XYZ* occupancy field. As can be seen, across all three degradation types, the proposed measure consistently obtains the highest sensitivity, outperforming the conventional occupancy-field metrics by a clear margin. This demonstrates that the proposed approach captures progressive structural degradation more effectively (See Supplementary materials for the exact values).

To further investigate this behavior, Figure 5 illustrates the progressive structural degradation of a representative occupancy field under increasing resolution reduction. The corresponding responses of the different quality measures are reported in Table 3. Unlike the image distortions considered in the previous subsection, all metrics preserve the correct ordering of the degradation levels, i.e., larger degradations consistently produce larger measured differences. However, correctly ordering degradations alone is insufficient if the response becomes increasingly compressed as degradation severity grows. In particular, volumetric IoU, one of the most widely adopted evaluation metrics for occupancy fields and INRs, exhibits a progressively attenuated response, with successive increases in degradation producing increasingly smaller relative changes in the metric. Similar behavior is observed for Chamfer distance and F-score, indicating that these measures increasingly underestimate the severity of additional structural degradation. A quantitative analysis of this attenuation is provided in the supplementary material. In contrast, **`SymKL`** maintains a strong and consistent response throughout the degradation sequence, preserving not only the correct ordering of progressively degraded occupancy fields but also clear separation between successive degradation levels. This experiment therefore reveals a complementary failure mode to that observed for natural images. Whereas image-quality metrics may fail by violating monotonicity under progressively stronger distortions, conventional occupancy-field metrics remain monotonic but exhibit an attenuated response that progressively underestimates additional degradation. The proposed approach addresses both behaviors by providing a stable, monotonic, and highly responsive characterization of degradation severity across occupancy fields.

Table 3: Responses of occupancy-field quality measures under increasing downsampling.

Metric	**DS1.5**	**DS2**	**DS3**	**DS4**	**DS6**
Chamfer ($\times 10^{-5}$)	2.910	4.072	5.118	9.062	17.860
F-score	0.959	0.932	0.893	0.759	0.581
IoU	0.960	0.945	0.933	0.905	0.866
`SymKL`	**0.180**	**0.367**	**0.505**	**1.287**	**2.953**

3.3 AUDIO

To further evaluate the proposed degradation measure across signal modalities, we next consider audio signals. We use the *dev-clean* split of the LibriSpeech dataset (Panayotov et al., 2015), from which we extract 54 reference clips of 3 seconds each (16 kHz) spanning 33 speakers. Each reference clip is progressively degraded using six common audio degradation types, additive Gaussian noise

with SNR levels of $\{30, 25, 20, 15, 10\}$ dB, sample dropout with dropout rates of $\{2, 5, 10, 20, 35\}\%$, high-pass filtering with normalized cutoff frequencies of $\{0.02, 0.05, 0.10, 0.20, 0.40\}$, spectral notching with normalized notch widths of $\{0.02, 0.05, 0.10, 0.20, 0.35\}$, clipping with peak-amplitude thresholds of $\{0.80, 0.50, 0.30, 0.15, 0.07\}$, and μ-law companding with quantization levels of $\{256, 128, 64, 32, 16\}$. This results in a total of 1,620 degraded audio samples. We compare the proposed measure against widely used full-reference audio quality metrics, including SNR, LSD, PESQ, and STOI.

The right figure of Figure 4 illustrates the normalized-mean relative sensitivity across all degradation levels for each degradation type. As can be seen **SymKL** obtains the highest sensitivity across all six degradations, demonstrating a stronger and more consistent response to changes in audio quality than the conventional audio-specific metrics. The largest gains are observed for spectral notching, clipping, and μ-law companding, while the proposed measure also attains the highest sensitivity for additive Gaussian noise, sample dropout, and high-pass filtering. These results indicate that **SymKL** provides a robust and modality-agnostic characterization of degradation that extends effectively from images and occupancy fields to audio signals.

4 Additional Results

The supplementary material provides additional quantitative analyses supporting the main paper, including all referenced quantitative comparisons, extended qualitative visualizations, and experiments on medical imaging datasets, offering a broader evaluation across diverse signal modalities.

5 Conclusion

We introduced **SymKL**, a modality-agnostic representation-space degradation measure that quantifies signal degradation through the Fisher information geometry of a frozen implicit neural representation (INR). Unlike conventional full-reference measures, which rely on modality-specific representations, **SymKL** compares reference and degraded signals through a common Fisher representation derived from a shared INR. Because the measure depends only on a frozen reference representation and the reconstruction objective, rather than the underlying corruption process or its severity, it applies unchanged across modalities and naturally extends to signals for which no established full-reference measure exists. To enable efficient computation, we derived an exact closed-form expression for the diagonal Fisher of SIREN networks. Across images, audio, and 3D occupancy fields, **SymKL** consistently tracks degradation severity, remaining monotonic and discriminative even where conventional measures saturate or violate monotonicity. More broadly, this work suggests that learned neural representations themselves can serve as the substrate for signal analysis. Just as pixels, waveforms, and point clouds have traditionally formed the basis for degradation measures, future signal-processing algorithms may instead operate directly on the information geometry of learned representations, providing unified tools that naturally extend across diverse signal modalities.

6 AI use statement

In this work, we used generative AI tools for grammar correction and writing enhancement, as well as to assist in generating some methodological figures. All AI-assisted text and figures were carefully reviewed and verified by the authors for accuracy and consistency with the proposed methodology. Generative AI was not used for research ideation, methodology development, data generation or analysis, experimental design, or the formulation of scientific claims and conclusions. We take full responsibility for the final content of this work, including text and figures produced with the assistance of generative AI.

References

Alessandro Achille, Michael Lam, Rahul Tewari, Avinash Ravichandran, Subhransu Maji, Charless C Fowlkes, Stefano Soatto, and Pietro Perona. Task2vec: Task embedding for meta-learning.

In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6430–6439, 2019.

Shun-Ichi Amari. Natural gradient works efficiently in learning. *Neural computation*, 10(2):251–276, 1998.

Awsaf. Coco 2017 dataset. `https://www.kaggle.com/datasets/awsaf49/coco-2017-dataset`, 2020. Kaggle dataset.

Vinod Kumar Chauhan, Jiandong Zhou, Ping Lu, Soheila Molaei, and David A Clifton. A brief review of hypernetworks in deep learning. *Artificial Intelligence Review*, 57(9):250, 2024.

Yinbo Chen and Xiaolong Wang. Transformers as meta-learners for implicit neural representations. In *European Conference on Computer Vision*, pp. 170–187. Springer, 2022.

Luca De Luigi, Adriano Cardace, Riccardo Spezialetti, Pierluigi Zama Ramirez, Samuele Salti, Luigi Di Stefano, et al. Deep learning on implicit neural representations of shapes. In *The Eleventh International Conference on Learning Representations*, volume 2023, pp. 1–39. ICLR, 2023.

Emilien Dupont, Hyunjik Kim, SM Eslami, Danilo Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one. *arXiv preprint arXiv:2201.12204*, 2022a.

Emilien Dupont, Hrushikesh Loya, Milad Alizadeh, Adam Goliński, Yee Whye Teh, and Arnaud Doucet. Coin++: Neural compression across modalities. *arXiv preprint arXiv:2201.12904*, 2022b.

Ziya Erkoç, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. Hyperdiffusion: Generating implicit neural fields with weight-space diffusion. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 14300–14310, 2023.

Amer Essakine, Yanqi Cheng, Chun-Wun Cheng, Lipei Zhang, Zhongying Deng, Lei Zhu, Carola-Bibiane Schönlieb, and Angelica I Aviles-Rivero. Where do we stand with implicit neural representations? a technical and performance survey. *arXiv preprint arXiv:2411.03688*, 2024.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pp. 1126–1135. PMLR, 2017.

Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. Dreamsim: Learning new dimensions of human visual similarity using synthetic data. *arXiv preprint arXiv:2306.09344*, 2023.

Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008.

Dhananjaya Jayasundara and Vishal M Patel. Implicit neural representations: A signal processing perspective. *arXiv preprint arXiv:2604.15047*, 2026.

Dhananjaya Jayasundara, Heng Zhao, Demetrio Labate, and Vishal M Patel. Mire: Matched implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8279–8288, 2025.

Don H Johnson. Signal-to-noise ratio. *Scholarpedia*, 1(12):2088, 2006.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Jonathan Le Roux, Scott Wisdom, Hakan Erdogan, and John R Hershey. Sdr–half-baked or well done? In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 626–630. IEEE, 2019.

Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1775–1787, 2023.

Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–3. IEEE, 2019.

Alexander Ly, Maarten Marsman, Josine Verhagen, Raoul PPP Grasman, and Eric-Jan Wagenmakers. A tutorial on fisher information. *Journal of Mathematical Psychology*, 80:40–55, 2017.

Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.

NIH Clinical Center. Nih chest x-ray dataset. `https://www.kaggle.com/datasets/nih-chest-xrays/data`, 2017. Accessed: 2026-02-26.

Jim Nilsson and Tomas Akenine-Möller. Understanding ssim. *arXiv preprint arXiv:2006.13846*, 2020.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *Transactions on Machine Learning Research Journal*, 2024.

Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 5206–5210. IEEE, 2015.

Tianming Qiu, Christos Sonis, and Hao Shen. Ensuring semantics in weights of implicit neural representations through the implicit function theorem. *arXiv preprint arXiv:2601.23181*, 2026.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Fiana Raiber and Oren Kurland. Kullback-leibler divergence revisited. In *Proceedings of the ACM SIGIR international conference on theory of information retrieval*, pp. 117–124, 2017.

Sameera Ramasinghe and Simon Lucey. Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In *European Conference on Computer Vision*, pp. 142–158. Springer, 2022.

Hippolyt Ritter, Aleksandar Botev, and David Barber. A scalable laplace approximation for neural networks. In *International conference on learning representations*, 2018.

Antony W Rix, John G Beerends, Michael P Hollier, and Andries P Hekstra. Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. In *2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221)*, volume 2, pp. 749–752. IEEE, 2001.

David M Rouse and Sheila S Hemami. Understanding and simplifying the structural similarity metric. In *2008 15th IEEE international conference on image processing*, pp. 1188–1191. IEEE, 2008.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk. Wire: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 18507–18516, 2023.

Arya Shah. Breast ultrasound images dataset. `https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset`, 2020. Accessed: 2026-02-26.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

Matthew Tancik, Ben Mildenhall, Terrance Wang, Divi Schmidt, Pratul P Srinivasan, Jonathan T Barron, and Ren Ng. Learned initializations for optimizing coordinate-based neural representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2846–2855, 2021.

Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3405–3414, 2019.

Tong Wu, Liang Pan, Junzhe Zhang, Tai Wang, Ziwei Liu, and Dahua Lin. Density-aware chamfer distance as a comprehensive metric for point cloud completion. *arXiv preprint arXiv:2111.12702*, 2021.

Zhuoqian Yang, Mathieu Salzmann, and Sabine Süsstrunk. Weight space representation learning with neural fields. *arXiv preprint arXiv:2512.01759*, 2025.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018.

Hui Zhang, Xueliang Zhang, and Guanglai Gao. Training supervised speech separation system to improve stoi and pesq directly. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5374–5378. IEEE, 2018a.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018b.

Dingfu Zhou, Jin Fang, Xibin Song, Chenye Guan, Junbo Yin, Yuchao Dai, and Ruigang Yang. Iou loss for 2d/3d object detection. In *2019 international conference on 3D vision (3DV)*, pp. 85–94. IEEE, 2019.

A Appendix

A.1 Related Works

Implicit neural representations. INRs model signals as continuous functions that map spatial or temporal coordinates to signal values. Their ability to represent complex, high-frequency content is enabled through techniques such as positional encodings or specialized activation functions, including periodic nonlinearities (Sitzmann et al., 2020) and space-frequency compact nonlinearities (Ramasinghe & Lucey, 2022; Saragadam et al., 2023; Jayasundara et al., 2025). Owing to their flexible coordinate-based formulation, INRs have been successfully applied to a wide range of signal modalities, including neural radiance fields, occupancy fields, signed-distance fields, and neural compression, where the signal is represented by the parameters of a neural network (Essakine et al., 2024; Jayasundara & Patel, 2026). Existing work primarily employs INRs as a representation for accurately reconstructing or compressing signals. In contrast, we use the learned INR as a representation to be analyzed rather than reconstructed, treating it as a common substrate on which a unified degradation measure can be defined across diverse signal modalities.

Learning in the weight space of neural representations. Recent work has increasingly explored learning directly from the parameters of neural representations rather than their outputs. Functa represents each fitted neural field by a learned modulation vector, which serves as a compact representation for downstream tasks (Dupont et al., 2022a), while inr2vec learns embeddings of INRs for shape understanding and retrieval (De Luigi et al., 2023). More generally, weight-space learning methods process neural network parameters directly to learn representations of trained models, enabling tasks such as classification, retrieval, and self-supervised representation learning (Qiu et al., 2026; Yang et al., 2025; Erkoç et al., 2023). In contrast, our objective is not to learn representations from network parameters. Instead, we use the Fisher profile of a single reference INR to characterize how degradation alters the functional organization of a learned neural representation, yielding a degradation measure rather than a learned embedding.

Fisher information as a descriptor of neural representations. The Fisher information matrix characterizes the sensitivity of a model's predictions to perturbations of its parameters and has become a fundamental tool in statistical learning and optimization. It forms the basis of natural-gradient optimization (Amari, 1998), underpins Laplace approximations for Bayesian neural networks (Ritter et al., 2018), and its diagonal has been widely adopted as a measure of parameter importance, most notably in elastic weight consolidation for continual learning (Kirkpatrick et al., 2017). Beyond optimization and uncertainty estimation, Fisher information has also been used as a representation of learned models. Closest to our work, Task2Vec represents learning tasks using the diagonal Fisher information of a fixed probe network after adaptation to each task, enabling task similarity to be quantified through comparisons of the resulting Fisher vectors (Achille et al., 2019). Our work builds on the same principle that parameter sensitivity reflects the information encoded by a neural network, but applies it in a fundamentally different setting. Rather than comparing datasets or learning tasks, we characterize degradation within a single signal using a fixed INR fitted to the pristine reference. The network parameters remain unchanged, and the Fisher profiles differ only through the target signal used for evaluation. Furthermore, we derive an exact closed-form expression for the diagonal Fisher information of SIREN, eliminating the expensive per-sample gradient computations required by conventional Fisher estimation.

Efficient INR fitting. Training an INR independently for every signal can be computationally expensive, motivating approaches that accelerate INR optimization. One direction employs hypernetworks to predict the parameters of a target network from an input representation (Chauhan et al., 2024). This idea has been extended to INRs using transformer-based weight decoders that directly predict INR parameters (Chen & Wang, 2022). A complementary direction is meta-learning, which seeks parameter initializations that can be rapidly adapted to new signals with only a few optimization steps (Finn et al., 2017). This mechanism has proven effective for INRs through learned initializations (Tancik et al., 2021) and modulation-based adaptation (Dupont et al., 2022b). Our approach combines these two paradigms by employing a hypernetwork conditioned on DINOv2 features (Oquab et al., 2024) to predict the weights of a SIREN, while meta-training the hypernetwork using a first-order MAML objective so that its predictions provide high-quality initializations for efficient test-time refinement.

A.2 HYPERNETWORK TRAINING DETAILS

Training data. The hypernetwork is trained on the union of MS-COCO (Awsaf, 2020) and LSDIR (Li et al., 2023), totaling approximately 203k images. LSDIR provides high-resolution photographic content that complements COCO by mitigating its bias toward small, object-centric images.

Stage 1: reconstruction pretraining. The hypernetwork is first trained to predict SIREN parameters that reconstruct the input image. For each image, we sample N_c coordinates, evaluate the predicted SIREN at those coordinates, and minimize the MSE between the predicted and corresponding ground-truth pixel values. Training follows the three-phase resolution curriculum summarized in Table 4. A low-resolution phase with large batches establishes the coordinate-to-parameter mapping, while two subsequent higher-resolution phases progressively adapt the model to native-resolution image statistics, with the learning rate reduced at each transition. The DINOv2 encoder is fine-tuned throughout using a learning rate equal to $0.1\times$ the base learning rate. We observed that encoder adaptation, rather than decoder capacity, is the primary contributor to performance: quadrupling the decoder improves reconstruction by only approximately 0.3 dB, whereas unfreezing the encoder provides a gain of 3–4 dB.

	Phase 1	Phase 2	Phase 3
Max side (px)	224	420	518
Learning rate	2×10^{-4}	1×10^{-4}	5×10^{-5}
Coords per image N_c	16,384	65,536	65,536
Initialised from	scratch	Phase 1	Phase 2

Table 4: Resolution curriculum for reconstruction pretraining. Validation PSNR is measured at each phase's own resolution and is not comparable across columns.

Stage 2: first-order meta-fine-tuning. The reconstruction objective in Stage 1 optimizes the quality of the hypernetwork's zero-shot prediction. However, at inference time the predicted SIREN parameters are used only as an initialization and are subsequently refined through per-image optimization. To better match this inference procedure, we perform a second training stage based on a first-order MAML (FOMAML) objective that directly optimizes reconstruction quality after refinement. For each training image (a 378^2 crop), the hypernetwork predicts an initial parameter vector $\theta_0 = h_\phi(x)$. These parameters are refined by $K = 32$ Adam updates with learning rate 10^{-4}, each computed on a fresh random sample of $65{,}536$ image coordinates, producing adapted parameters θ_K. The post-adaptation reconstruction loss $\mathcal{L}_{\text{post}} = \frac{1}{|c|}\|f_{\theta_K}(c) - y\|^2$ is then evaluated on another independently sampled coordinate set. Ideally, the hypernetwork would be updated with the gradient $\partial\mathcal{L}_{\text{post}}/\partial\theta_0$, but computing it exactly requires differentiating through the entire inner optimization via $\partial\theta_K/\partial\theta_0$. We instead adopt the standard first-order MAML approximation $\partial\theta_K/\partial\theta_0 \approx I$, under which $\partial\mathcal{L}_{\text{post}}/\partial\theta_0 \approx \nabla_{\theta_K}\mathcal{L}_{\text{post}}$. This gradient is planted onto the prediction through the surrogate objective

$$\mathcal{L}_{\text{meta}} = \Big\langle \theta_0, \underbrace{\nabla_{\theta_K}\mathcal{L}_{\text{post}}}_{\text{detached}} \Big\rangle + \lambda_{\text{zs}}\,\mathcal{L}_{\text{zs}}, \qquad \lambda_{\text{zs}} = 0.1,$$

whose gradient with respect to θ_0 is by construction exactly the detached vector, so backpropagating through $\theta_0 = h_\phi(x)$ delivers the approximate meta-gradient to ϕ. The auxiliary zero-shot loss $\mathcal{L}_{\text{zs}} = \frac{1}{|c'|}\|f_{\theta_0}(c') - y\|^2$ prevents the quality of the initialization itself from degrading during meta-training. Additional details are provided in Table 5.

A.3 DERIVATION

We derive the diagonal empirical Fisher for a general coordinate MLP with elementwise differentiable activations and then specialize the result to the sinusoidal activations used in SIREN. Consider an L-layer coordinate MLP defined as $h_n^{(0)} = u_n, \qquad z_n^{(\ell)} = W_\ell h_n^{(\ell-1)} + b_\ell, \qquad h_n^{(\ell)} = \phi_\ell\Big(z_n^{(\ell)}\Big), \quad \ell < L$, with a linear output $f_\theta(u_n) = z_n^{(L)} \in \mathbb{R}^C$. For coordinate u_n, we use the squared reconstruction loss, $\mathcal{L}_n(\theta) = \frac{1}{2C}\,\|f_\theta(u_n) - y(u_n)\|_2^2$. .Let $\delta_n^{(\ell)} := \frac{\partial\mathcal{L}_n}{\partial z_n^{(\ell)}} \in \mathbb{R}^{d_\ell}$, denote

Component	Configuration	Params
Encoder	DINOv2 ViT-B/14	87.4M
Decoder	8 Trans blks, d=1024, 16 hds	100.8M
Query tokens	1,027 learned queries	1.1M
Prediction heads	Per-neuron linear heads	1.1M
Base SIREN	Learnable initialization	0.2M
Total		190.5M
Target SIREN	5 layers, 256 hidden units, $\omega_0 = 30$	198,915

Table 5: Hypernetwork architecture and parameter counts. The hypernetwork predicts the parameters of a 5-layer SIREN using 1,027 learned query tokens, corresponding to its hidden and output neurons.

the pre-activation error at layer ℓ, and let $a_n^{(\ell)} := h_n^{(\ell-1)}$ denote the corresponding layer input. As $z_n^{(\ell)} = W_\ell a_n^{(\ell)} + b_\ell$,, the gradient of the coordinate loss with respect to the weight matrix is $\frac{\partial \mathcal{L}_n}{\partial W_\ell} = \delta_n^{(\ell)} \big(a_n^{(\ell)}\big)^\top$. Therefore, the gradient of each layer is a rank-one outer product. Elementwise, this can be written as $\frac{\partial \mathcal{L}_n}{\partial [W_\ell]_{ij}} = \delta_{n,i}^{(\ell)} a_{n,j}^{(\ell)}$. Similarly, the bias gradient is $\frac{\partial \mathcal{L}_n}{\partial b_\ell} = \delta_n^{(\ell)}$. The diagonal empirical Fisher is defined as the average of the elementwise squared per-coordinate gradients, $F = \frac{1}{N}\sum_{n=1}^N \nabla_\theta \mathcal{L}_n \odot \nabla_\theta \mathcal{L}_n$. For an individual weight $[W_\ell]_{ij}$, its Fisher entry is therefore

$$
\begin{aligned}
[F_{W_\ell}]_{ij} &= \frac{1}{N}\sum_{n=1}^N \left(\frac{\partial \mathcal{L}_n}{\partial [W_\ell]_{ij}}\right)^2 \\
&= \frac{1}{N}\sum_{n=1}^N \big(\delta_{n,i}^{(\ell)}\big)^2 \big(a_{n,j}^{(\ell)}\big)^2 .
\end{aligned} \tag{1}
$$

Let $\Delta^{(\ell)} \in \mathbb{R}^{N\times d_\ell}$ stack the error vectors $\delta_n^{(\ell)}$ row-wise, and let $A^{(\ell)} \in \mathbb{R}^{N\times d_{\ell-1}}$ stack the layer inputs $a_n^{(\ell)}$ row-wise. The sum in Eq. equation 1 can then be evaluated for all parameters of the layer using a single matrix multiplication:

$$
F_{W_\ell} = \frac{1}{N}\left(\Delta^{(\ell)} \odot \Delta^{(\ell)}\right)^\top \left(A^{(\ell)} \odot A^{(\ell)}\right). \tag{2}
$$

The corresponding bias entries are

$$
F_{b_\ell} = \frac{1}{N}\sum_{n=1}^N \delta_n^{(\ell)} \odot \delta_n^{(\ell)}. \tag{3}
$$

The error vectors required in Eqs. equation 2 and equation 3 are obtained using the standard backward recursion. As the output layer is linear, the output error is $\delta_n^{(L)} = \frac{1}{C}\left(f_\theta(u_n) - y(u_n)\right)$. For the preceding layers, applying the chain rule gives $\delta_n^{(\ell-1)} = \left(W_\ell^\top \delta_n^{(\ell)}\right) \odot \phi'_{\ell-1}\left(z_n^{(\ell-1)}\right), \quad 2 \le \ell \le L$,. For the SIREN used in this work, the activation and its derivative are $\phi_\ell(z) = \sin(\omega_\ell z), \quad \phi'_\ell(z) = \omega_\ell \cos(\omega_\ell z)$. The backward recursion therefore becomes $\delta_n^{(\ell-1)} = \left(W_\ell^\top \delta_n^{(\ell)}\right) \odot \omega_{\ell-1} \cos\left(\omega_{\ell-1} z_n^{(\ell-1)}\right)$. The sinusoidal activation enters the computation only through this derivative factor; the remaining derivation applies to any elementwise differentiable activation.

Computing the diagonal empirical Fisher requires one forward pass to obtain the activations, one backward recursion to obtain the error vectors, and one matrix accumulation per layer. Its arithmetic complexity is $\mathcal{O}\Big(N\sum_{\ell=1}^L d_\ell d_{\ell-1}\Big) = \mathcal{O}(NP)$, where P is the total number of network parameters. This has the same asymptotic complexity as a standard full-batch loss-and-gradient evaluation.

Modality	PSNR	SSIM	MS-SSIM	SymKL
MRI	0.420	0.327	0.408	**0.516**
Ultrasound	1.062	0.246	0.408	**1.479**
X-ray	0.995	0.734	0.891	**1.729**

Table 6: Mean relative sensitivity (elasticity) across three medical imaging modalities. Higher values indicate greater sensitivity to degradation. **SymKL** consistently achieves the highest sensitivity across all modalities.

Metric	$\times 2$	$\times 4$	$\times 8$	$\times 16$
PSNR	27.652	23.278	22.219	21.886
SSIM	0.894	0.813	0.784	0.775
MS-SSIM	0.914	0.841	0.821	0.821
SymKL	**0.589**	**2.063**	**2.700**	**2.926**

Table 7: Metric values for the MRI example shown in Fig. 6. While conventional metrics saturate at higher undersampling factors, **SymKL** continues to respond to increasing degradation.

A.4 Additional Results

A.4.1 Medical Imaging

Dataset Preparation: We evaluate **SymKL** across ultrasound, MRI, and X-ray under controlled degradation protocols.
Ultrasound: We use the Breast Ultrasound Images Dataset (Shah, 2020). Speckle noise is synthetically injected with severity levels $L \in \{1, 2, 4, 8, 16\}$ using a multiplicative model. To simulate restoration pipelines, we apply median filtering. Each despeckled image is compared against the given reference. **MRI:** We use accelerated reconstructions from fastMRI (Zbontar et al., 2018), with undersampling factors $k \in \{2, 4, 8, 16\}$. Each undersampled reconstruction is evaluated relative to its fully sampled reference to assess degradation under increasing sampling sparsity. **X-ray:** Chest radiographs from the NIH dataset (NIH Clinical Center, 2017) are degraded using signal-dependent Poisson noise, where λ is proportional to dose levels $\{1.0, 0.7, 0.5, 0.35, 0.25\}$. Each simulated low-dose image is compared to its clean counterpart.
Sensitivity Analysis: For all experiments, we employ a five-layer sinusoidal INR with 300 hidden neurons per layer. We compare **SymKL** against commonly used similarity measures, including PSNR, SSIM, and MS-SSIM. Table 6 reports the average relative sensitivities of each metric with respect to its corresponding reference. It is worth noting that learned perceptual metrics such as LPIPS, CLIP, and DreamSim are not suitable for this analysis, as they are predominantly trained on RGB images. Although they can produce similarity scores for other modalities through a forward pass, these scores are generally unreliable unless the models are specifically adapted to the target modality, which is often impractical, particularly in the medical imaging domain. In contrast, **SymKL** is modality-agnostic and can be applied directly without any modality-specific adaptation.

Across all evaluated medical modalities, **SymKL** consistently achieves the highest sensitivity. This property is particularly critical in medical imaging, where small degradations may conceal subtle anatomical structures, distort lesion boundaries, or alter clinically relevant biomarkers. A similarity metric with high progressive sensitivity ensures that incremental quality deterioration is reliably detected, supporting safer reconstruction, compression, and acquisition pipelines. These results indicate that widely used similarity metrics are insufficiently responsive in medical settings, whereas **SymKL** provides a more reliable representation-level assessment of degradation severity.

Figure 6 and Table 7 present a representative MRI example under progressively increasing undersampling factors. From $\times 8$ to $\times 16$, the undersampling severity doubles, yet MS-SSIM changes by only 0.0008 ($0.8213 \rightarrow 0.8205$), indicating that it largely saturated. In contrast, **SymKL** continues to increase ($2.70 \rightarrow 2.93$), reflecting the increased degradation. Notably, all metrics remain monotonic for this example; the limitation of the conventional metrics is therefore not incorrect ordering, but their inability to meaningfully respond to increasingly severe degradation, consistent with the aggregate MRI sensitivity analysis.

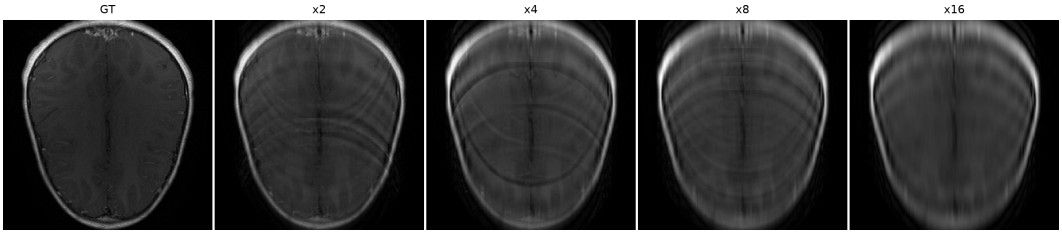

Figure 6: Representative MRI reconstruction under progressively increasing undersampling factors ($\times 2$, $\times 4$, $\times 8$, and $\times 16$). Although the visual degradation continues to increase with undersampling severity, conventional similarity metrics become increasingly insensitive at high degradation levels, whereas **SymKL** continues to reflect the progression of degradation.

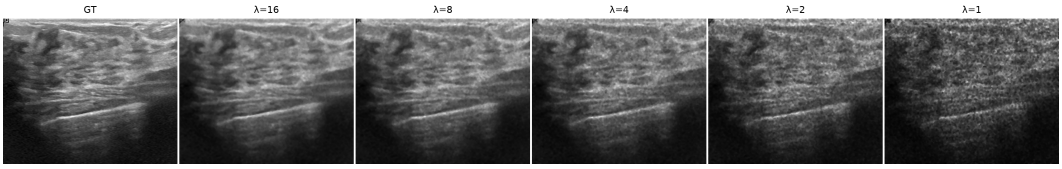

Figure 7: Representative ultrasound reconstruction under progressively decreasing photon counts ($\lambda = 16, 8, 4, 2, 1$). As the photon count decreases, speckle noise increases substantially. While conventional similarity metrics exhibit limited dynamic range, **SymKL** closely follows the increasing degradation severity.

Figure 7 and Table 8 present a representative ultrasound example with progressively decreasing photon count ($\lambda = 16, 8, 4, 2, 1$), corresponding to increasing speckle noise severity. As the photon count is halved, the physical noise power approximately doubles, producing a clear progression in visual degradation. **SymKL** closely follows this trend, increasing from 0.339 to 0.807, 1.924, 4.922, and 14.418, corresponding to an overall increase of approximately $42.6\times$. In contrast, conventional similarity metrics exhibit comparatively limited dynamic range; for example, SSIM changes only from 0.787 to 0.500 despite the $16\times$ change in photon count.

A.4.2 RGB IMAGING

Table 9 summarizes additional distortion types commonly encountered in practical image processing. As shown in the table, **SymKL** consistently achieves the highest overall sensitivity across all distortion types.

Figure 8 and Table 10 present a representative RGB example with progressively stronger color quantization (64, 48, 32, 16, and 8 colors). The visual degradation increases monotonically as the number of available colors decreases. However, several existing metrics exhibit little response or even slight inversions at mild distortion levels. For example, DreamSim remains nearly unchanged from 0.012 to 0.011 to 0.011 as the palette is reduced from 64 to 32 colors, while SSIM, MS-SSIM, and LPIPS all slightly improve from Q1 to Q2 despite the increased degradation. In contrast, **SymKL** increases consistently from 1.323 to 1.708 to 2.423, and reaches 10.942 at the highest distortion level, corresponding to an overall increase of approximately $8.3\times$. This example illustrates that **SymKL** remains responsive to progressively increasing degradation, whereas existing metrics can become insensitive or even locally non-monotonic.

Figure 9 and Table 11 present a representative example from the KADID-10k quantization distortion under progressively increasing severity levels. As the number of quantization levels decreases from 20 to 7, the visual degradation becomes increasingly pronounced. While DreamSim remains monotonic, SSIM, LPIPS, and CLIP exhibit local inversions despite the increasing distortion. In contrast, **SymKL** increases consistently from 0.068 to 0.138, 0.232, 0.491, and 1.125, providing a smooth and progressively stronger response to the increasing degradation. This example demonstrates that

Metric	$\lambda = 16$	$\lambda = 8$	$\lambda = 4$	$\lambda = 2$	$\lambda = 1$
PSNR	29.083	27.331	24.877	21.620	17.376
SSIM	0.787	0.746	0.690	0.607	0.500
MS-SSIM	0.926	0.903	0.868	0.809	0.719
SymKL	**0.339**	**0.807**	**1.924**	**4.922**	**14.418**

Table 8: Metric values for the ultrasound example shown in Fig. 7. As the photon count decreases, **SymKL** closely tracks the increasing speckle noise, whereas conventional metrics exhibit a comparatively limited response.

Table 9: Mean relative sensitivity on additional KADID-10k distortions image-domain and weight-space baselines. Best results are shown in bold and second-best results are underlined. Higher is better.

	Image-domain baselines						**Weight-space baselines**					
Distortion	**PSNR**	**SSIM**	**MS-SSIM**	**LPIPS**	**CLIP**	**DSim**	**F-Cos**	**F-ℓ_2**	**G-Cos**	**G-ℓ_2**	**Adam-Cos**	**SymKL**
Color quant.	0.969	0.393	0.646	0.579	0.551	0.924	0.433	1.120	0.244	0.835	0.100	**1.240**
JPEG	0.803	0.515	0.855	0.925	0.850	0.822	0.378	0.985	0.323	0.743	0.108	**1.435**
Color noise	0.951	0.334	0.596	0.528	0.288	0.511	0.447	0.958	0.323	0.484	0.091	**1.253**
Mult. noise	0.953	0.380	0.601	0.602	0.291	0.550	0.753	0.964	0.420	0.494	0.113	**1.200**
Denoising	0.639	0.361	0.565	0.427	0.451	0.603	0.306	0.709	0.097	0.504	0.057	**0.820**

Metric	Q1 (64)	Q2 (48)	Q3 (32)	Q4 (16)	Q5 (8)
PSNR	26.066	25.290	23.985	21.828	18.394
SSIM	0.832	0.833	0.803	0.750	0.662
MS-SSIM	0.975	0.977	0.972	0.957	0.930
LPIPS	0.089	0.080	0.103	0.169	0.284
CLIP	0.990	0.968	0.984	0.980	0.955
DreamSim	0.012	0.011	0.011	0.021	0.108
SymKL	**1.323**	**1.708**	**2.423**	**4.586**	**10.942**

Table 10: Metric values for the RGB example shown in Fig. 8. Existing metrics exhibit limited sensitivity and occasional local inversions under mild color quantization, whereas **SymKL** responds consistently to increasing degradation.

Metric	Q1 (20)	Q2 (16)	Q3 (13)	Q4 (10)	Q5 (7)
PSNR	34.129	32.228	30.882	28.303	25.308
SSIM	0.840	0.813	0.795	0.800	0.750
MS-SSIM	0.960	0.950	0.945	0.936	0.930
LPIPS	0.199	0.251	0.232	0.278	0.225
CLIP	0.945	0.923	0.934	0.885	0.902
DreamSim	0.026	0.034	0.046	0.054	0.055
SymKL	**0.068**	**0.138**	**0.232**	**0.491**	**1.125**

Table 11: Metric values for the KADID-10k quantization example shown in Fig. 9. While several metrics exhibit local inversions under increasing quantization severity, **SymKL** responds monotonically and with substantially greater sensitivity.

SymKL preserves both monotonicity and sensitivity, whereas several existing perceptual metrics fail to consistently reflect the progression of quantization distortion.

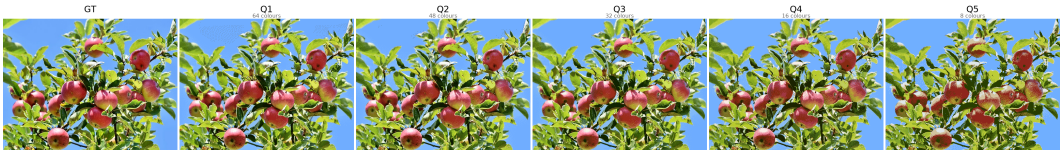

Figure 8: Representative RGB image under progressively increasing color quantization (64, 48, 32, 16, and 8 colors). Although the visual degradation increases monotonically as the color palette is reduced, several existing perceptual metrics exhibit limited sensitivity or slight inversions at lower distortion levels. In contrast, **SymKL** consistently reflects the increasing degradation severity.

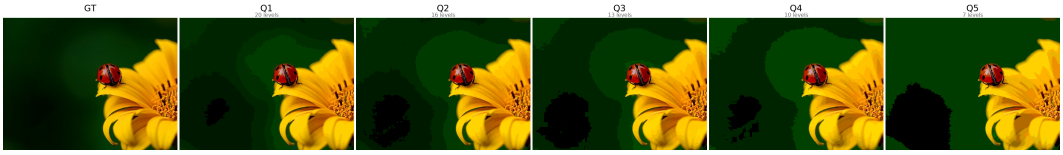

Figure 9: Representative example from the KADID-10k quantization distortion under progressively increasing severity levels. As the number of quantization levels decreases, the visual degradation becomes increasingly pronounced. While several existing perceptual metrics exhibit local inversions, **SymKL** consistently tracks the progression of degradation.

A.4.3 Occupancy Fields

As described in the main paper, **SymKL** is further evaluated on implicit 3D occupancy fields using diverse degradations. Table 12 reports the corresponding unnormalized sensitivity values for the experiment presented in Fig. 4(a) of the main paper.

Additionally, we evalute **SymKL** on Thai Statue occupancy field, increasing morphological erosion produces a monotonic increase in geometric degradation through inward displacement of the object surface while preserving the overall topology. Figure 10 illustrates the degradation sequence, and Table 13 summarizes the corresponding measures.

F-score. The F-score measures the fraction of surface points that lie within a fixed matching threshold ($\tau = 2$ voxels). Once the surface displacement exceeds this threshold, no points satisfy the matching criterion and the score saturates at zero. In this experiment, saturation occurs from erosion level 3 onward, although the surface continues to move substantially. Consequently, multiple visually distinct degradation levels receive the same score. This behavior is reflected in Table 13, where the F-score exhibits high sensitivity only before saturation and zero sensitivity thereafter.

Volumetric IoU. Unlike the F-score, volumetric IoU remains monotonic throughout the degradation sequence. However, because it measures overlap between occupied volumes rather than surface displacement, erosion removes only a relatively thin boundary shell while most of the interior remains unchanged. As a result, IoU decreases only gradually despite increasingly severe geometric degradation, leading to consistently sub-proportional sensitivity ($S < 1$) across all degradation levels.

Weight-space sensitivity. In contrast, **SymKL** remains monotonic throughout the entire erosion sequence and achieves the highest relative sensitivity (Table 13). Unlike the F-score, it continues to distinguish degradation levels after the threshold-based matching has saturated, and unlike volumetric IoU, its response is not attenuated by the large unchanged interior volume. Consequently, it provides a more consistent measure of progressively increasing geometric degradation across the full erosion range.

A.4.4 Audio

Table 12: Sensitivity (S) on occupancy field degradations, reported as the mean relative sensitivity over all valid degradation steps. Higher is better.

Degradation	Chamfer	1-Fscore	1-IoU	SymKL
Blur	4.657	5.126	2.130	**12.378**
Erode	2.715	1.463	0.876	**2.914**
Down-res	1.492	2.087	0.903	**2.777**

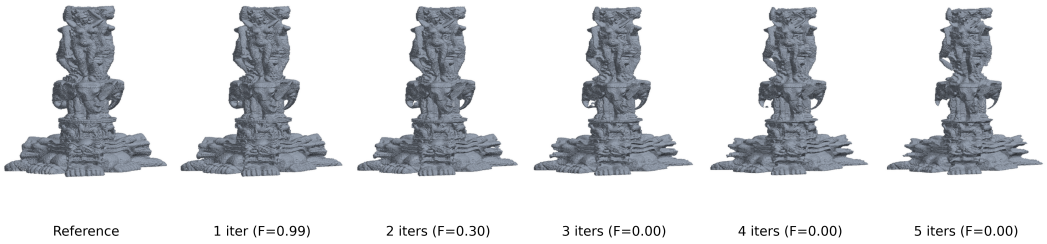

Figure 10: Progressively eroded occupancy fields of the Thai statue. Increasing erosion depth produces progressively larger geometric deviations through inward displacement of the object surface. While the geometric degradation remains visually distinguishable throughout the sequence, conventional geometric similarity measures become progressively less responsive, whereas **SymKL** continues to distinguish successive degradation levels.

Table 14: Sensitivity (S) on audio degradations, reported as the mean across clips of the mean over degradation steps. Higher is better.

Degradation	SNR	LSD	PESQ	STOI	SymKL
Gaussian Noise	2.778	0.214	0.225	1.828	**2.831**
Dropout	1.539	0.742	0.403	1.161	**1.878**
High-pass	1.290	0.592	0.805	1.134	**1.310**
Notch	3.473	0.590	2.178	1.722	**3.918**
Clipping	3.244	2.207	2.951	2.699	**3.590**
Mu-law	3.244	0.347	1.103	2.393	**3.404**

Table 14 shows the unnormalized metric values corresponding to Fig. 4(b) in the main paper.

Metric	L1	L2	L3	L4	L5
F-score	0.901	0.213	**0.000**	**0.000**	**0.000**
IoU	0.932	0.869	0.809	0.753	0.701
SymKL	**0.841**	**3.778**	**8.690**	**15.441**	**23.848**

Table 13: Metric values on the progressively eroded Thai statue occupancy field. F-score saturates after level 3 and IoU decreases gradually, whereas **SymKL** continues to distinguish increasingly severe geometric degradation.

Figure 11 and Table 15 present a representative audio example corrupted by progressively increasing additive Gaussian noise. As the SNR decreases from 30 dB to 10 dB, the degradation becomes increasingly apparent in both the waveform and the perceptual quality. While PESQ quickly approaches its lower operating range ($1.245 \rightarrow 1.109 \rightarrow 1.068 \rightarrow 1.055 \rightarrow 1.049$), providing only limited discrimination between the more severe degradation levels, **SymKL** continues to increase substantially ($1.169 \rightarrow 4.178 \rightarrow 13.419 \rightarrow 44.197 \rightarrow 138.400$), corresponding to an overall increase of approximately $118\times$. This example demonstrates that **SymKL** remains responsive over the full degradation range, whereas conventional perceptual measures rapidly lose dynamic range.

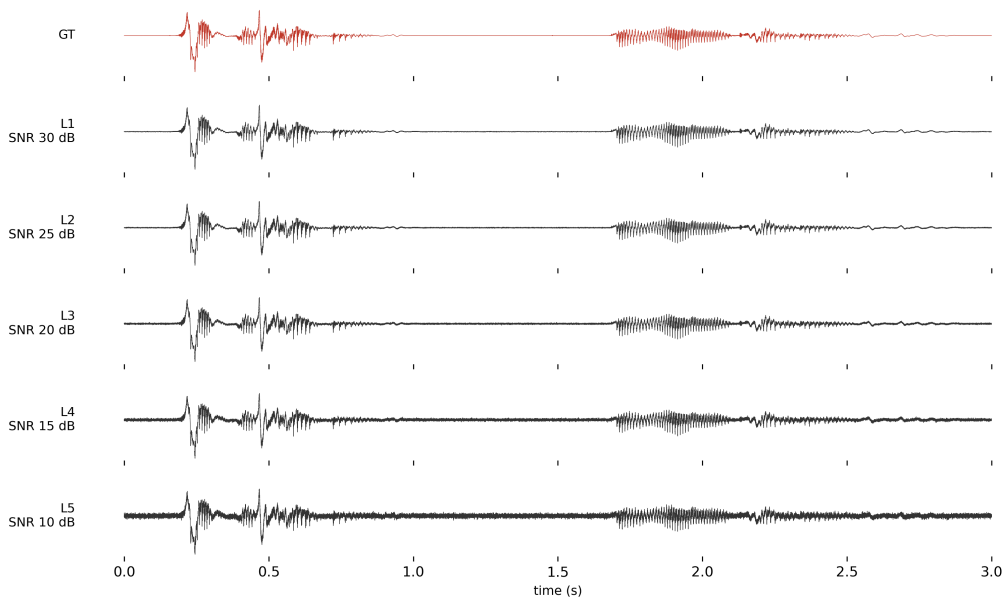

Figure 11: Representative audio waveform under progressively increasing additive Gaussian noise (30, 25, 20, 15, and 10 dB SNR). As the SNR decreases, the waveform becomes increasingly corrupted. While conventional perceptual metrics rapidly lose dynamic range at higher noise levels, **SymKL** continues to distinguish successive degradation levels.

Metric	L1 (30 dB)	L2 (25 dB)	L3 (20 dB)	L4 (15 dB)	L5 (10 dB)
SNR (dB)	29.992	24.999	20.051	14.953	9.992
LSD	27.694	31.738	35.891	40.312	44.740
PESQ	1.245	1.109	1.068	1.055	1.049
STOI	0.975	0.951	0.924	0.889	0.842
SymKL	**1.169**	**4.178**	**13.419**	**44.197**	**138.400**

Table 15: Metric values for the audio example shown in Fig. 11. As the noise level increases, PESQ rapidly approaches its lower operating range, whereas **SymKL** continues to respond strongly across the entire degradation sequence.

A.5 Experimental Setup

Training setup. All experiments are conducted on a server equipped with eight NVIDIA RTX A5000 GPUs (24 GB each), dual AMD EPYC 7352 24-Core processors (48 CPU cores), and 503 GB of system memory.