# Anisotropic Fourier Features for Neural Image-Based Rendering and Relighting

**Huangjie Yu**[1,3,4], **Anpei Chen**[1], **Xin Chen**[1], **Lan Xu**[1], **Ziyu Shao**[1], **Jingyi Yu**[1,2]

[1] School of Information Science and Technology, ShanghaiTech University
[2] Shanghai Engineering Research Center of Intelligent Vision and Imaging
[3] University of Chinese Academy of Sciences
[4] Shanghai Institute of Microsystem and Information Technology
{yuhj, chenap, chenxin2, xulan1, shaozy, yujingyi}@shanghaitech.edu.cn

## Abstract

Recent neural rendering techniques have greatly benefited image-based modeling and relighting tasks. They provide a continuous, compact, and parallelable representation by modeling the plenoptic function as multilayer perceptrons (MLPs). However, vanilla MLPs suffer from spectral biases on multidimensional datasets. Recent rescues based on isotropic Fourier features mapping mitigate the problem but still fall short of handling heterogeneity across different dimensions, causing imbalanced regression and visual artifacts such as excessive blurs. We present an anisotropic random Fourier features (RFF) mapping scheme to tackle spectral biases. We first analyze the influence of bandwidth from a different perspective: we show that the optimal bandwidth exhibits strong correlations with the frequency spectrum of the training data across various dimensions. We then introduce an anisotropic feature mapping scheme with multiple bandwidths to model the multidimensional signal characteristics. We further propose an efficient bandwidth searching scheme through iterative golden-section search that can significantly reduce the training overload from polynomial time to logarithm. Our anisotropic scheme directly applies to neural surface light-field rendering and image-based relighting. Comprehensive experiments show that our scheme can more faithfully model lighting conditions and object features as well as preserve fine texture details and smooth view transitions even when angular and spatial samples are highly imbalanced.

## Introduction

There has been renewed interest in conducting image-based modeling and rendering (IBMR), such as new view synthesis and post-capture relighting, largely due to recent advances in neural modeling and rendering. Applications range from surface reflectance modeling to photorealistic relighting in virtual reality and augmented reality. While traditional approaches rely on ray interpolation and tedious hand-crafted blending schemes, recent neural advances implicitly encode continuous surface attributes like reflectance or appearance. Such implicit solutions enable 1) powerful interpolation capabilities with continuous representation in both spatial and angular domains; 2) de-factor compression scheme; for example, neural relighting compresses light transport matrices by a factor of 200; 3) parallel implementation capability using modern computing hardware to achieve near real-time performance.

The core of these implicit image-based rendering techniques is the adoption of deep fully-connected networks (also called multilayer perceptrons or MLPs). Such coordinate-based MLPs take low-dimensional coordinates as inputs and compactly regress a continuous output for various attributes like shape, density, or color (Mescheder et al. 2019; Chen et al. 2018; Mildenhall et al. 2020; Yariv et al. 2020). Theoretically, the MLPs can approximate any measurable functions to any desired degree of accuracy (Hornik, Stinchcombe, and White 1989). However, it is known that standard MLPs tend to learn low-frequency functions due to "spectral bias" (Rahaman et al. 2019), leading to severe blur artifact for neural image-based rendering and relighting.

To overcome such drawbacks of vanilla MLPs, previous researchers (Mildenhall et al. 2020; Zhong et al. 2019) experimentally adopt a heuristic sinusoidal mapping of the input coordinates called positional encoding to recover high-frequency details in protein structures and surface textures. At the same time, the recent work (Sitzmann et al. 2020) utilizes the "SIREN" network to substitute the traditional ReLU activation function in MLPs with a periodic one. The most notable method (Tancik et al. 2020) adopts Gaussian random Fourier features (RFF), which relies on a carefully chosen standard deviation (bandwidth) to sample frequencies from an isotropic normal distribution and generate sinusoids for feature transformation. However, current methods still utilize a time-consuming grid-search scheme for the bandwidth parameter, leading to inefficient training overload. Moreover, such isotropic Gaussian RFF mapping treats every coordinate equally, without considering the heterogeneity of a plenoptic function across various dimensions for image-based rendering and relighting. Thus it suffers from imbalanced regression and the loss of high-frequency details (over-smoothness) or rough signal estimation (under-smoothness). For example, a surface light field displays much more substantial color variation among adjacent surface points than among different viewing angles. However, only using a single bandwidth fails to balance different degrees of smoothness in the spatial and angular domain.

In this paper, we attack the above challenges and present an anisotropic RFF mapping scheme for a range of neural

implicit image-based rendering and relighting tasks. Our approach significantly improves the performance of MLPs for these IBMR applications by taking the RFF mapping into the new anisotropic realm.

Specifically, we first show that applying Gaussian RFF mapping results in a kernelized neural network by composing the neural tangent kernel (NTK) (Jacot, Gabriel, and Hongler 2018) of the vanilla MLP with a radial basis function (RBF) kernel, where the bandwidth of the RBF kernel plays a crucial role in controlling the smoothness of MLPs. Then through empirical analysis, we find that the optimal bandwidth is determined by the density of training data and the signal's frequency spectrum. Based on such critical insight, we propose an anisotropic feature mapping that employs multiple bandwidths to match plenoptic characteristics in different dimensions for the image-based rendering and relighting tasks. To enable automatic bandwidth searching and efficient training, we propose an iterative algorithm based on linear golden-section search (GSS) to search multiple bandwidth parameters. Our method significantly outperforms the computationally expensive brute-force grid search algorithm by reducing the number of searches from polynomial time to logarithm time during training. We apply our novel anisotropic scheme for various IBMR tasks. For neural surface light field rendering, we implement a 2-bandwidth Gaussian kernel and the corresponding anisotropic RFF mapping to model both the high-frequency texture of the object surfaces and the low color variation in the angular domain simultaneously. We adopt a similar implementation for neural image relighting to capture the reflectance variation in spatial and angular domains. Our approach faithfully models lighting conditions and object features to recover fine texture details and preserve smooth viewing transition. To summarize, our main contributions include:

- We introduce an anisotropic Fourier features mapping scheme for multidimensional representations of neural IBMR tasks where the traditional isotropic mapping method fails. The key motivation is that the optimal bandwidth is determined by the training data density and signal's frequency spectrum in each dimension of plenoptic functions through empirical analysis about the isotropic Fourier features.
- We propose an efficient golden-section-based algorithm for searching optimal bandwidths under multidimensional settings, which reduces training time remarkably.
- We achieve high-quality results on both neural surface light field rendering and relighting tasks, which outperform previous state-of-the-art methods favorably.

## Related Work

Our work is mainly related to recent advances in deep network-based neural rendering (NR) as well as emerging machine learning techniques for further refining NR schemes.

**Neural Rendering and Relighting.** There have been significant progresses on applying implicit neural representations (Park et al. 2019; Henzler, Mitra, and Ritschel 2020; Yariv et al. 2020; Niemeyer et al. 2020) for image-based modeling and rendering (Zhang et al. 2021; Chen et al. 2018; Mildenhall et al. 2020). Different from traditional IBMR techniques that aim to conduct view interpolation from input images with hand-crafted blending schemes (Hedman et al. 2016; Penner and Zhang 2017; Buehler et al. 2001), neural rendering schemes set out to learn such schemes. Hedman et al. proposed to learn the blending weights using a convolutional neural network (Hedman et al. 2018) whereas Chen et al. utilized a fully-connected neural network to represent a surface light field to encode geometry and appearance simultaneously for efficient interpolation. However, these schemes still exhibit excessive blurs (Chen et al. 2018), as we will demonstrate, due to the misleading similarity measure inherent to standard MLPs. Alternatively, Aliev et al. presented a point-based neural renderer for modeling geometry and colors. The seminal work of Neural Radiance Field (NeRF) (Mildenhall et al. 2020; Martin-Brualla et al. 2021) adopted a learnable volumetric representation that consists of opacity and color at each position along every ray and applied an MLP-based network to recover the complete plenoptic function.

Another important IBMR problem is image-based relighting stemming from the pioneer LightStage system (Debevec et al. 2000). A LightStage captures the reflectance of objects (e.g., human faces) under a dense set of directional lights to form the light transport matrix. These images can then be used for relighting human faces, e.g., to match a new environment in virtual production. The majority of the efforts have been focused on reducing data size (Reddy, Ramamoorthi, and Curless 2012; Wang et al. 2009; Ren et al. 2013). Ren et al. proposed an MLP-based method to reconstruct the light transport matrix from a sparse set of reflectance images (Ren et al. 2015). Specifically, they store an average reflectance map in advance to tackle over-smoothed outputs of neural networks. Likewise, Xu et al. leveraged a CNN-based method to exploit coherence in the light transport function (Xu et al. 2018).

**High-Frequency Functions Learning.** Our goal is to study how to enhance MLP-based NR schemes by reducing blurring and aliasing artifacts. Recent methods (Rahaman et al. 2019; Basri et al. 2020) have shown that MLPs with a ReLU activation function fail to adequately represent fine details in complex low-dimensional signals due to the spectral bias. These studies reveal that standard MLPs have difficulties in learning high-frequency functions. To address this problem, existing works either replace the traditional ReLU activation function with periodic ones (Sitzmann et al. 2020) or lift the input coordinates into a Fourier feature space (Mildenhall et al. 2020; Tancik et al. 2020). With the latter approach, original inputs are mapped through a series of sinusoidal functions with either fixed (Mildenhall et al. 2020) or isotropic Gaussian-sampled frequencies (Tancik et al. 2020). By adopting the Neural Tangent Kernel (NTK) theory, Tancik et al. show that coordinated-based MLPs correspond to kernels with a rapid spectrum falloff and subsequently cause extremely slow convergence

to high-frequency components. Although the application of isotropic Fourier features mitigates the problem, it does not consider anisotropy across dimensions. In particular, we show that existing isotropic Fourier features fall short under multidimensional regression tasks with non-uniform data distribution or different changing rate of signal values. In IBMR tasks, they lead to visual artifacts such as loss of details or specularity.

## Preliminaries

**Isotropic Fourier Features Mapping (IFFM).** To overcome the spectral bias introduced by standard MLPs, an isotropic Fourier features mapping scheme was introduced by (Tancik et al. 2020). It maps input coordinates $\boldsymbol{x}$ into the following function $\phi(\boldsymbol{x})$ before passing them into an MLP:

$$\phi(\boldsymbol{x}) := \tfrac{1}{\sqrt{m}}[\cos(\boldsymbol{w}_1^T\boldsymbol{x}+\boldsymbol{b}_1),\ldots,\cos(\boldsymbol{w}_m^T\boldsymbol{x}+\boldsymbol{b}_m)], \quad (1)$$

where $\boldsymbol{x}$ is the original input; $b_i$ is a random number in $[0, 2\pi)$. All components of $\boldsymbol{w}_i$ are random frequencies drawn from a uni-variate normal distribution.

By Bochner's theorem (Liu et al. 2020), we obtain the following asymptotic results:

$$\lim_{m\to\infty}\phi(\boldsymbol{x})\phi(\boldsymbol{x'})^T = \exp(-\frac{1}{2}\|\frac{\boldsymbol{x}-\boldsymbol{x'}}{\alpha}\|^2), \quad (2)$$

where $\alpha$ is the reciprocal of the standard deviation of frequencies in $\phi(\boldsymbol{x})$. In our IBMR tasks, $\alpha$ denotes the smoothing parameter that controls reconstruction bandwidth.

**Perspective From Neural Tangent Kernel.** The neural tangent kernel (NTK) theory (Lee et al. 2019; Jacot, Gabriel, and Hongler 2018) shows that a trained neural network in the infinite-width regime equates to a Gaussian process, therefore making the dynamics of the training algorithm tractable and amenable to quantitative analysis. Specifically, the NTK theory states that given an MLP after sufficient iterations of training, for every test input point $\boldsymbol{x}$, the output of such an MLP denoted by $\hat{f}(\boldsymbol{x})$ converges to a normal distribution as the network width goes to infinity. The asymptotic distribution of $\hat{f}(\boldsymbol{x})$ is given by

$$\mathcal{N}\left(k_g(\boldsymbol{x},\mathcal{X})K^{-1}\mathcal{Y}, k_g(\boldsymbol{x},\boldsymbol{x}) - k_g(\boldsymbol{x},\mathcal{X})K^{-1}k_g(\mathcal{X},\boldsymbol{x})\right), \quad (3)$$

where the function $k_g(\cdot,\cdot)$ is defined over data pairs to measure the similarity of two inputs. It is called the "neural tangent kernel" inherent to an MLP of a certain architecture. $\mathbf{K}$ is an $n \times n$ matrix with entries $\mathbf{K}_{ij} = k_g(\boldsymbol{x}_i, \boldsymbol{x}_j)$. It is worth noting that $k_g(\boldsymbol{x}, \boldsymbol{x'})$ is an *increasing* function with respect to the dot product of $\boldsymbol{x}$ and $\boldsymbol{x'}$ (Lee et al. 2018, 2019). Therefore, applying IFFM to the input of an MLP results in a composed neural tangent kernel in place of $k_g(\boldsymbol{x}, \boldsymbol{x'})$:

$$\tilde{k}_g(\boldsymbol{x},\boldsymbol{x'}) := k_g(\phi(\boldsymbol{x})\phi(\boldsymbol{x'})^T) \approx k_g(\exp(-\frac{1}{2}\|\frac{\boldsymbol{x}-\boldsymbol{x'}}{\alpha}\|^2)). \quad (4)$$

In this way, the original dot product NTK is replaced with an bell-shaped kernel. In other words, the MLP is *kernelized* by an RBF kernel after applying IFFM to its input. We also explore scenarios where we make use of Laplacian (and other) kernels. More details can be found in Appendix A.

Now we focus on the MLPs kernelized by the RBF kernel. Intuitively, after kernelization, the prediction at a test point $\boldsymbol{x}$ is a weighted average of the $n$ training points with blending weights positively correlated to the values of $\tilde{k}_g(\boldsymbol{x}_i - \boldsymbol{x})$. Thus, the bandwidth parameter $\alpha$ tunes MLPs to lie between over-smoothing and under-smoothing regimes by adjusting blending schemes. Thus, selecting appropriate $\alpha$ is crucial for improving the performance of neural image-based rendering and relighting.

## Our Method

### Drawback of IFFM

In this subsection, we conduct numerical analysis and derive the critical observation: the optimal bandwidth is determined by the density of training data and the signal's frequency spectrum.

For basic numerical setting, we follow Tancik et al. and generate ground truth 1D functions "$1/f^{\lambda}$" on the interval $[0, 1)$: sample a vector from a standard Gaussian with size 2000, multiply its $i^{th}$ entry with $1/i^{\lambda}$ and return the real components of its inverse Fourier transform. The performance metric to measure the goodness of a bandwidth is defined by the mean integrated squared error (MISE):

$$\begin{aligned}\mathrm{MISE}(\hat{f}(x;\alpha)) &= \int_0^1 \mathrm{E}\left(\hat{f}(x;\alpha) - f(x)\right)^2 dx \\ &= \int_0^1 \mathrm{Var}\left(\hat{f}(x;\alpha)\right) dx + \int_0^1 \left(\mathrm{E}\,\hat{f}(x;\alpha) - f(x)\right)^2 dx,\end{aligned} \quad (5)$$

where $f(x)$ denotes the ground-truth signal. The expectation inside the integral is taken with respect to network initialization. We reinitialize and train the same network several times and take the average in practice. The variance and expectation of $\hat{f}(x;\alpha)$ are given in Eqn.3.

In Fig. 1, we show how the generalization error (MISE) depends on the bandwidth under various data density and signal frequencies. In Fig. 1a, we train MLPs to fit a "$1/f^{1.5}$" signal using IFFM ($m = 256$) under various bandwidth settings, which reveals that smaller bandwidth (less smoothing) is required as more training points are included, i.e., higher data density. This relation is more evident in Fig. 1b, where we show bandwidths that minimize generalization error versus data density under different signal frequencies. In Fig. 1c, we train MLPs to fit signals of various frequencies given the same number of training points. It shows that when the signal has more features (smaller $\lambda$), a smaller bandwidth is called for faithful reconstruction. In contrast, smoother signals correspond to larger bandwidths. Fig. 1d also illustrates this trend under various training settings.

We have an intuitive explanation for the above numerical results. Since $\mathrm{Var}\left(\hat{f}(x;\alpha)\right)$ in Eqn.5 decreases as more training points are considered (Williams and Vivarelli 2000), we can obtain the upper bound of the integrated variance

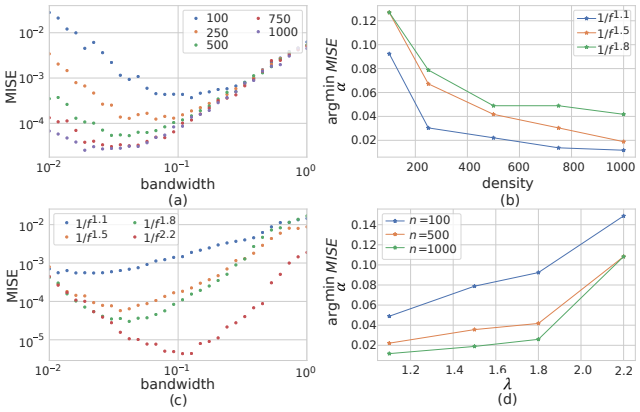


Figure 1: The selection of the bandwidth is a trade-off and relies on the number of training points per unit (a) and signal frequencies (c). Smaller bandwidths are called for when training points increase (b) or the signal shows stronger variations and more features (d).

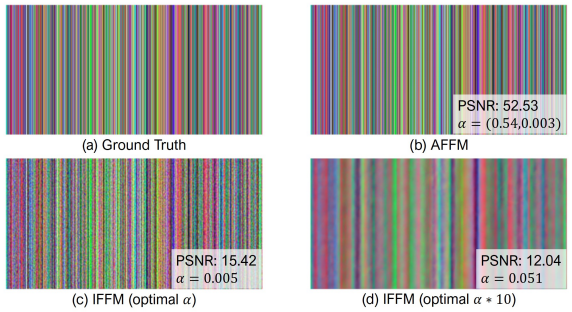


Figure 2: We compare the performance of IFFM and AFFM on regressing a stripe pattern image (a). Each column of the pattern is random color, making the pattern exhibit strong variation on the width dimension and none variation on the height dimension. The application of IFFM fails to regress such an image even with optimal bandwidth (c). The prediction is noisy due to the high variance caused by a small bandwidth. However, a larger bandwidth can lead to high bias (d) and over-smoothed prediction. AFFM alleviates this difficulty by introducing two bandwidths to capture diverse signal characteristics and outperforms IFFM remarkably (b).

term in Eqn.5 by considering 1 closest training point. The final upper bound is of order $k'_g(1)(n\alpha)^{-2}$ (More details can be found in Appendix B). This upper bound indicates that the integrated variance term depends on the network architecture and the training data density. It also indicates that a small $\alpha$ value will cause a high variance, which means each independently trained network will produce diverse outputs at the same point. This usually results in a spiky (under-smoothed) function estimate. Furthermore, we note that the squared bias term depends on the first and higher-order derivatives of the actual signal. The bias is significant when the absolute values of these derivatives are large. This usually happens in regions such as boundaries and features. In these cases, a small bandwidth is called for to reconstruct details. The dilemma of reducing variance or bias indicates that optimal bandwidth selection is a trade-off between the training data density and signal frequencies.

## Anisotropic Fourier Features Mapping (AFFM)

We know that IFFM with a single bandwidth fails to regress signals that show imbalanced characteristics (frequency and density) in different dimensions. This motivates us to develop a new mapping method for multidimensional tasks, especially for neural IBMR applications, where either the training data have different densities or the signal shows distinct smoothness across spatial and angular dimensions. To the best of our knowledge, we are the first to propose an anisotropic Fourier features mapping (AFFM) $\phi(x)$ in the field of MLPs:

$$\lim_{m\to\infty} \phi(\boldsymbol{x})\phi(\boldsymbol{x'})^T = \exp(-\frac{1}{2}\sum_i \|\frac{\boldsymbol{x}_i - \boldsymbol{x}'_i}{\alpha_i}\|^2), \qquad (6)$$

where the parameter $\alpha_i$ denotes the bandwidth for $i$-th dimension. It differs from IFFM in that we incorporate multiple bandwidths to match signal heterogeneity in different dimensions. It is worth noting that a group of dimensions can share the same bandwidth if they have similar semantic meanings. Basically, a large bandwidth corresponds to the dimension where the total curvature (variation) of the signal is small or the number of training points per unit is small. The construction of AFFM is also straightforward. In practice, we sample frequencies $\boldsymbol{w}_i$ from a multivariate normal distribution whose covariance is a diagonal matrix whose $i^{th}$ diagonal entry is $1/\alpha_i$.

To illustrate the superiority of our AFFM over IFFM, we train MLPs with both schemes to fit a strip-like square pattern using half pixels. Corresponding results are shown in Fig. 2, where the reconstructed image using AFFM is visually lossless and has a much higher PSNR than using IFFM. More specifically, Fig. 2(b) shows the results by applying AFFM, with two smoothing parameters for width and height dimensions, respectively. The bandwidth corresponding to the height dimension is significant to produce enough smoothness on that dimension, while the substantial variation on the width dimension is also faithfully reconstructed.

In contrast, applying IFFM to regress this image results in either noisy estimates (Fig. 2c) or loss of details (Fig. 2d). Fig. 2c shows the predicted result by applying IFFM with optimal bandwidth (highest PSNR). We observe that the optimal $\alpha$ value is small to compensate for a high variation on the width dimension. However, such small $\alpha$ value results in high output variance (large Var $(\hat{f}(x;\alpha))$) and causes noisy artifacts on the height dimension. In Fig. 2d, we employ a bandwidth that is 10 times larger than the optimal one. Nevertheless, it leads to losing high frequencies details in the width dimension by introducing more biases. Note that we are unable to find a good middle point.

## Efficient Search of (Near) Optimal Bandwidth

In implementing AFFM for practice, it is critical to search for (near) optimal bandwidth(s). The brute force method that

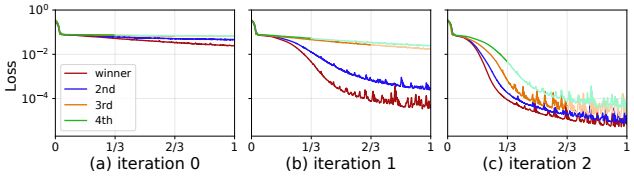


Figure 3: We show the first 3 iterations of a 2D GSS and corresponding training losses of the 4 candidates in each iteration. The most promising bandwidth candidate (red curve) usually prevails early in selection.

uses grid search to enumerate possible solutions is computationally prohibitive, leading to extremely high training overload. To tackle such a challenge, we adopt a linear golden-section search (GSS) algorithm as the start point. It is well known that GSS has several advantages over other search algorithms: 1) it is gradient-free; calculating gradients with respect to bandwidth is even more time-consuming; 2) it avoids the selection of initial point; 3) it saves function evaluations per iteration since each evaluation of bandwidth takes an entire network training; 4) it is natural to extend GSS to multidimensional scenarios.

**Uni-bandwidth Selection.** For uni-bandwidth setting, a linear GSS works as follows: given the initial interval inside which the optimal bandwidth lies, it first evaluates two bandwidth candidates that are located at proportional positions $0.382$ and $0.618$; then it reduces the interval around the minimum by a factor of $0.618$ (the golden ratio); it repeats until convergence. Note that one evaluation at each iteration is saved from the last iteration. Therefore, after $k$ iterations, the width of the search interval becomes $(0.618)^k$, while the number of network trainings is $k+1$. In this way, golden-section search reduces the number of searches from polynomial to logarithmic.

**Multi-bandwidth Selection.** The extension of linear GSS to multiple dimensions is natural. We provide an example in the two-bandwidth setting. A 2D GSS first selects two parameters in both dimensions, producing $4$ pairs of bandwidth candidates; then shrinks the search box around the minimum by a factor of $0.618^2$ and repeats until convergence. After $k$ iterations, we obtain a parameter box of area $0.618^{2k}$. The number of network training is $3k+1$. We provide the algorithm pseudocode of 2D GSS in Appendix C.

**Heuristics to Speed Up Selection.** We observe that in each iteration, the most promising pair of bandwidths tend to prevail early in the process. For example, in Figure 3, we plot 3 search iterations under the strip pattern regression discussed in the previous subsection. The most promising candidate (denoted by the red curve) always performs better than other candidates during the whole training process. This motivates us to throw away 'bad' candidates sooner progressively. A similar idea is proposed in (Li et al. 2017). Specifically, we divide each iteration into 3 phases. We take out one-third of the total candidates at the end of the first two phases. At the end of the final phase, we take out all but the winning candidate. We empirically find that $5$ iterations are sufficient to achieve good practice convergence.

## AFFM for Neural IBMR Tasks

Neural implicit IBMR applications usually rely on plenoptic functions, which usually behaves differently across various dimensions, i.e., spatial and angular domains, for generating high-quality and photorealistic results. Thus we apply our AFFM scheme to two typical MLP-based IBMR scenarios, i.e., surface light field (SLF) rendering and image-based relighting.

An SLF represents the radiance of rays emitting from all points on the surface of an object to all directions and hence serves as an image-based representation of the plenoptic function (Chen et al. 2018). Specifically, a neural SLF defines a continuous mapping using a coordinated-based MLP:

$$L : M \times \Omega \to \text{RGB}, \tag{7}$$

where $M \subseteq \mathbf{R}^3$ denotes the set of points on the surface (spatial) and $\Omega$ denotes all unit vectors in $\mathbf{R}^3$ (angular). For a fixed surface point $m \in M$, the slice $L(m, \cdot)$ represents outgoing radiances of the point to all directions. If the surface is Lambertian, $L(m, \cdot)$ is constant. For non-Lambertian surface points, $L(m, \cdot)$ exhibits view dependency properties such as specularity. Nevertheless, such variation in the angular domain is generally much smaller than texture variations in the spatial domain as illustrated in (Chen et al. 2018). Therefore, IFFM with a global bandwidth fails to address this heterogeneity, as shown in the next section. To this end, for SLF rendering, we apply AFFM using a two-bandwidth Gaussian kernel and the corresponding anisotropic Fourier features. Our anisotropic scheme encodes both the high-frequency texture of the object surfaces in the spatial domain and the low color variation in the angular domain simultaneously.

Another representative IBMR problem is the image-based relighting stemming from the pioneer Light Stage system (Debevec et al. 2000). Such a technique aims to render photorealistic objects illuminated under novel lighting conditions from a fixed viewpoint with coordinate-based implicit OLAT (one light at a time) imagesets. According to the additivity of light, the relighting at a pixel $p$ is formulated as an integration on the sphere with respect to the angular lighting vector $\omega$:

$$L(p) = \int_{\Omega} R(\omega, p) E(\omega)\, d\omega \approx \frac{1}{4\pi N} \sum_{i=1}^{N} R(\omega_i, p) E(\omega_i), \tag{8}$$

where $R(\omega, p)$ is the reflectance at pixel $p$ in the direction $\omega$; $E(\omega)$ is the novel illumination (usually encoded in an environment map). Here $L$ denotes the relighting function. Usually, $R(\omega, p)$ shows more data coherence in the angular domain than in the spatial domain (Ren et al. 2015), since the BRDF of most materials is smooth with respect to incident light direction. Thus, we utilize MLPs with two-bandwidth AFFM to regress a continuous form of $R(\omega, p)$. The training dataset consists of discrete images to capture the reflectance variations in both spatial and angular domains. During the


Figure 4: The high-quality neural rendering results using our anisotropic random Fourier features mapping scheme for both the surface light field rendering and image-based relighting applications.

| | Synthetic ↑ | | | | | Real ↑ | |
|---|---|---|---|---|---|---|---|
| | Ship | Chicken | Lucy | Materials | Lego | Tang San Cai | Beer Can |
| No Mapping | 25.72/0.698 | 22.82/0.491 | 19.79/0.432 | 20.66/0.707 | 23.30/0.655 | 22.21/0.577 | 19.62/0.463 |
| IFFM | 26.74/0.732 | 24.09/0.573 | 21.75/0.600 | 23.39/0.832 | 25.10/0.745 | 26.33/0.782 | 22.38/0.689 |
| D-SLF | 26.12/0.710 | 25.61/0.761 | 21.78/0.586 | 22.53/0.794 | 27.54/0.860 | 24.08/0.719 | 22.53/0.746 |
| Ours | **28.12/0.806** | **27.36/0.803** | **28.76/0.880** | **26.87/0.913** | **31.23/0.931** | **29.61/0.903** | **25.94/0.864** |

Table 1: Quantitative comparisons on surface light-field rendering against various methods in terms of PSNR and SSIM. Our anisotropic scheme achieves consistently better results for various test scenes and metrics.

relighting phase after training, we sample a set of novel lights, query the MLP network for reflectances at each pixel to each light direction, and evaluate Eqn. 8 using the Monte Carlo approximation.

## Experiments

In this section, we qualitatively and quantitatively evaluate our anisotropic approach for two typical MLP-based IBMR tasks (surface light field rendering and image-based relighting), followed by the evaluation of our GSS-based iterative bandwidth searching scheme. We run our experiments on a PC with 2.2 GHz Intel Xeon 4210 CPU 64GB RAM and Nvidia TITAN RTX GPU. For all the applications, we use a ReLU fully-connected network with 8 layers and 256 channels in different solutions. The length of features is 8192 for SLF and 1024 for relighting. All the models are trained using the Adam (Kingma and Ba 2014) optimizer with a learning rate of $10^{-4}$ using PyTorch (Paszke et al. 2019). As shown in Fig. 4, our approach generates high-quality results for various MLP-based neural rendering and relighting tasks. It faithfully recovers texture details and preserves smooth viewing transition under challenging heterogeneous input manifolds.

### Comparison

Here we first compare our approach under the neural surface light field rendering task, which maps the spatial and angular coordinates of a surface to the RGB texture output. Let **AFFM** and **IFFM** denote our approach with anisotropic Fourier features mapping and the isotropic one from previous method, respectively. For thorough evaluation, we also compare against the state-of-the-art deep surface light field framework (Chen et al. 2018) denoted as **D-SLF**, and the one using standard MLP without feature mapping denoted as **No Mapping**. For quantitative evaluation, we adopt the peak signal-to-noise ratio (**PSNR**) and structural similarity index (**SSIM**) as metrics similar to previous methods. As shown in Tab. 1, our approach consistently outperforms the other baselines in terms of all these metrics under both real and synthetic datasets. The corresponding qualitative comparison in Fig. 5 also illustrates our capability to recover finely detailed texture results with rich view-dependent effects.

We further evaluate our approach under the image-based relighting task, which aims to render objects into novel lighting conditions from a fixed viewpoint with neural implicit OLAT imagesets. Similarly, we adopt the same PSNR metric and compare against the two IFFM and No Mapping baselines. As shown in Fig. 6 and Tab. 2, our anisotropic ap-

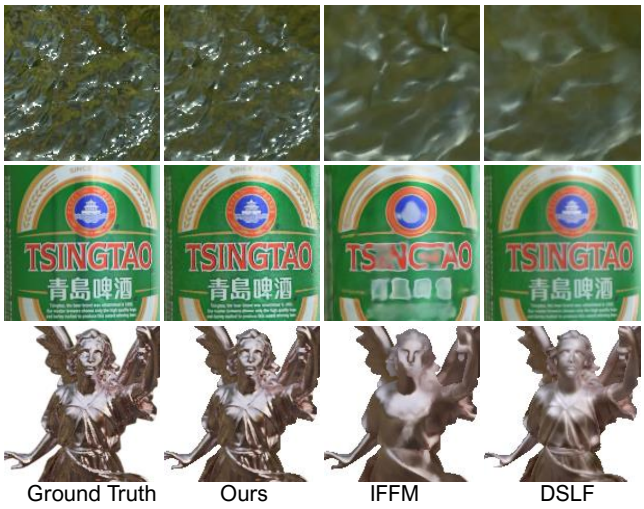


Figure 5: Qualitative comparisons on surface light field rendering. Our approach recovers fine detailed reflection and texture results remarkably across all the scenes.

| | No Mapping | IFFM | Ours |
|---|---|---|---|
| Armor (Eucalyptus) | 27.57 | 27.38 | **29.23** |
| Knight | 27.55 | 27.92 | **29.73** |
| Armor (Cathedral) | 27.06 | 27.16 | **28.03** |
| Knight | 29.68 | 30.06 | **32.72** |

Table 2: Quantitative comparisons on neural image-based relighting in terms of PSNR.

proach recovers high quality relighting details remarkably.

All these comparisons on various neural representation tasks illustrate that our approach with anisotropic Fourier features can faithfully model lighting conditions and object features under challenging heterogeneous input manifolds. Impressively, our approach renews the capability of popular MLP-based implicit representation for challenging IBMR tasks.

## Evaluation of Bandwidth Searching

We evaluate our GSS-based bandwidth searching scheme using the relighting task with 253 reflectance images of the "kneeling knight" imageset. Let **2D GSS** and **Linear GSS** denote the variations by applying our GSS-based searching scheme to the AFFM and IFFM methods, respectively, while **Grid search** denotes the original scheme in (Tancik et al. 2020). As shown in the training curves from Fig. 7, our GSS-based searching scheme for the bandwidth parameters consistently reduces the training overload and achieves roughly 7× speed-up for the training process compared to the brute-force grid searching. Besides, the superiority of our AFFM against the enhanced IFFM with GSS-based searching further demonstrates the effectiveness of our anisotropic Fourier features mapping design.

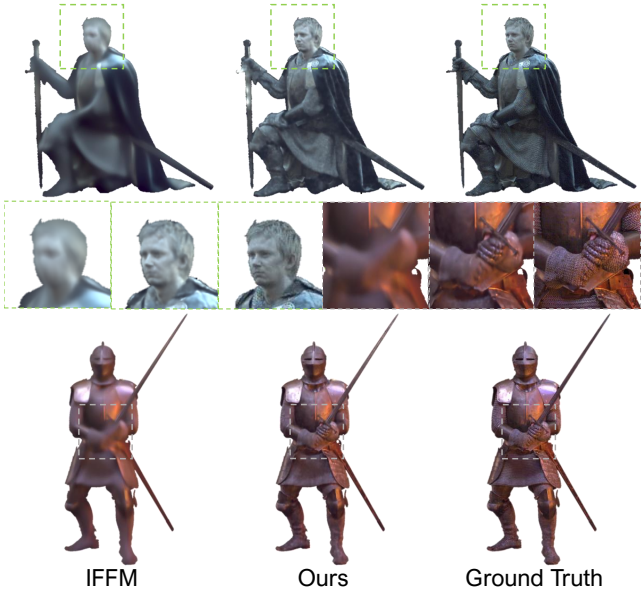


Figure 6: Qualitative comparisons on neural relighting task. Our approach recovers high quality relighting details.

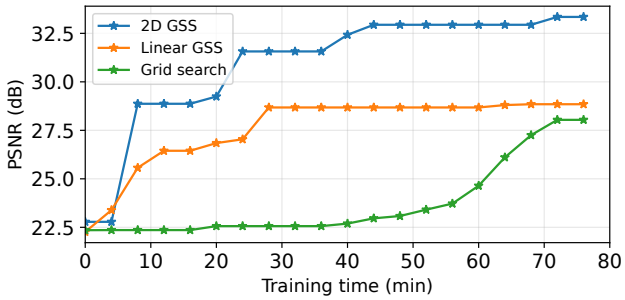


Figure 7: Evaluation of our GSS-baded bandwidth searching scheme. We provide numerical training time curves for three methods under the neural relighting task.

## Conclusions

In this paper, we presented a novel approach to take the isotropic Fourier features into the anisotropic realm for neural imaged-based rendering and relighting tasks. From the perspective of NTK theory, we connect MLPs using FFM with Gaussian processes. Then we empirically show the drawback of IFFM and propose AFFM. It employs multiple bandwidths to model signal heterogeneity under different dimensions, while our golden-section-based bandwidth searching scheme further reduces the training overload. Impressively, our anisotropic approach achieves state-of-the-art performance for both the neural SLF rendering and relighting. It successfully regresses fine texture details and preserves smooth viewing transition under challenging input manifolds. We believe that it is a critical step for understanding coordinate-based MLPs, with burgeoning applications using implicit neural representations for IBMR.

## Acknowledgements

This work was supported by NSFC programs (61976138, 61977047), the National Key Research and Development Program (2018YFB2100500), STCSM (2015F0203-000-06), and SHMEC (2019-01-07-00-01-E00003).

## References

Aliev, K.-A.; Sevastopolsky, A.; Kolos, M.; Ulyanov, D.; and Lempitsky, V. 2020. Neural Point-Based Graphics. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J.-M., eds., *Computer Vision – ECCV 2020*, 696–712. Cham: Springer International Publishing. ISBN 978-3-030-58542-6.

Basri, R.; Galun, M.; Geifman, A.; Jacobs, D.; Kasten, Y.; and Kritchman, S. 2020. Frequency bias in neural networks for input of non-uniform density. In *International Conference on Machine Learning*, 685–694. PMLR.

Buehler, C.; Bosse, M.; McMillan, L.; Gortler, S.; and Cohen, M. 2001. Unstructured Lumigraph Rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, 425–432. New York, NY, USA: Association for Computing Machinery. ISBN 158113374X.

Chen, A.; Wu, M.; Zhang, Y.; Li, N.; Lu, J.; Gao, S.; and Yu, J. 2018. Deep surface light fields. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 1(1): 1–17.

Debevec, P.; Hawkins, T.; Tchou, C.; Duiker, H.-P.; Sarokin, W.; and Sagar, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 145–156.

Hedman, P.; Philip, J.; Price, T.; Frahm, J.-M.; Drettakis, G.; and Brostow, G. 2018. Deep Blending for Free-Viewpoint Image-Based Rendering. *ACM Trans. Graph.*, 37(6).

Hedman, P.; Ritschel, T.; Drettakis, G.; and Brostow, G. 2016. Scalable Inside-out Image-Based Rendering. *ACM Trans. Graph.*, 35(6).

Henzler, P.; Mitra, N. J.; and Ritschel, T. 2020. Learning a neural 3d texture space from 2d exemplars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8356–8364.

Hornik, K.; Stinchcombe, M.; and White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5): 359–366.

Jacot, A.; Gabriel, F.; and Hongler, C. 2018. Neural tangent kernel: Convergence and generalization in neural networks. *arXiv preprint arXiv:1806.07572*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lee, J.; Sohl-dickstein, J.; Pennington, J.; Novak, R.; Schoenholz, S.; and Bahri, Y. 2018. Deep Neural Networks as Gaussian Processes. In *International Conference on Learning Representations*.

Lee, J.; Xiao, L.; Schoenholz, S.; Bahri, Y.; Novak, R.; Sohl-Dickstein, J.; and Pennington, J. 2019. Wide neural networks of any depth evolve as linear models under gradient descent. *Advances in neural information processing systems*, 32: 8572–8583.

Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; and Talwalkar, A. 2017. Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1): 6765–6816.

Liu, F.; Huang, X.; Chen, Y.; and Suykens, J. A. 2020. Random Features for Kernel Approximation: A Survey in Algorithms, Theory, and Beyond. *arXiv preprint arXiv:2004.11154*.

Martin-Brualla, R.; Radwan, N.; Sajjadi, M. S.; Barron, J. T.; Dosovitskiy, A.; and Duckworth, D. 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7210–7219.

Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; and Geiger, A. 2019. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4460–4470.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

Niemeyer, M.; Mescheder, L.; Oechsle, M.; and Geiger, A. 2020. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3504–3515.

Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 165–174.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, 8026–8037.

Penner, E.; and Zhang, L. 2017. Soft 3D Reconstruction for View Synthesis. *ACM Trans. Graph.*, 36(6).

Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; and Courville, A. 2019. On the spectral bias of neural networks. In *International Conference on Machine Learning*, 5301–5310. PMLR.

Reddy, D.; Ramamoorthi, R.; and Curless, B. 2012. Frequency-space decomposition and acquisition of light transport under spatially varying illumination. In *European Conference on Computer Vision*, 596–610. Springer.

Ren, P.; Dong, Y.; Lin, S.; Tong, X.; and Guo, B. 2015. Image based relighting using neural networks. *ACM Transactions on Graphics (ToG)*, 34(4): 1–12.

Ren, P.; Wang, J.; Gong, M.; Lin, S.; Tong, X.; and Guo, B. 2013. Global illumination with radiance regression functions. *ACM Transactions on Graphics (TOG)*, 32(4): 1–12.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33.

Tancik, M.; Srinivasan, P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.; and Ng, R. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33.

Wang, J.; Dong, Y.; Tong, X.; Lin, Z.; and Guo, B. 2009. Kernel Nyström method for light transport. In *ACM SIGGRAPH 2009 papers*, 1–10.

Williams, C. K.; and Vivarelli, F. 2000. Upper and lower bounds on the learning curve for Gaussian processes. *Machine Learning*, 40(1): 77–102.

Xu, Z.; Sunkavalli, K.; Hadap, S.; and Ramamoorthi, R. 2018. Deep Image-Based Relighting from Optimal Sparse Samples. *ACM Trans. Graph.*, 37(4).

Yariv, L.; Kasten, Y.; Moran, D.; Galun, M.; Atzmon, M.; Ronen, B.; and Lipman, Y. 2020. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33.

Zhang, J.; Liu, X.; Ye, X.; Zhao, F.; Zhang, Y.; Wu, M.; Zhang, Y.; Xu, L.; and Yu, J. 2021. Editable Free-viewpoint Video Using a Layered Neural Representation. *arXiv preprint arXiv:2104.14786*.

Zhong, E. D.; Bepler, T.; Davis, J. H.; and Berger, B. 2019. Reconstructing continuous distributions of 3D protein structure from cryo-EM images. *arXiv preprint arXiv:1909.05215*.