

SCALING LAWS AND COMPLEXITY OF GENERATIVE MODELS: A MULTIFRACTAL PERSPECTIVE

Anonymous authors

Paper under double-blind review

ABSTRACT

Assessing the functional aspects of a generative model (GM) is crucial to technological advancement. However, existing evaluation metrics are often insensitive to distributional changes and rarely correlate with perceptual fidelity. In addition, their oversimplified assumptions limit their ability to assess morphological fidelity and contextuality in application areas such as medical imaging and industrial machine vision. Hence, domain-agnostic, robust, and reliable GM evaluation remains an unresolved problem in generative AI and is an ongoing research paradigm. So, this work introduces the concept of multifractality, a scaling technique adapted from statistical physics for GM evaluation. Several multifractal markers are proposed as new metrics to analyze the scaling behavior of GMs. They characterize the structural complexity of long-range correlation patterns in GM-generated images. Non-parametric statistics-based hypothesis testing is formulated to assess the disparity in morphological organization between synthesized and actual data. These metrics are extensively validated using benchmark GMs on real-world datasets. Furthermore, multifractal spectrum analysis provides deeper insights into a GM’s complexity origin and plausible spectral bias explainability.

1 INTRODUCTION

Contemporary GMs, such as *generative adversarial networks* (GANs) [Goodfellow et al. \(2020\)](#); [Karras et al. \(2019; 2020; 2021\)](#), *variational autoencoders* (VAEs) [Kingma & Welling \(2013\)](#); [Razavi et al. \(2019\)](#), *normalizing flows* (NF) [Kingma & Dhariwal \(2018\)](#), and *diffusion models* [Ho et al. \(2020b\)](#); [Song et al. \(2020\)](#); [Kingma et al. \(2021\)](#); [Song et al. \(2021\)](#); [Dhariwal & Nichol \(2021\)](#); [Vahdat et al. \(2021\)](#) have demonstrated unprecedented fidelity, often exceeding human evaluative capacities [Dhariwal & Nichol \(2021\)](#); [Rombach et al. \(2022\)](#). Consequently, this has raised questions on the sufficiency of existing evaluation metrics to the extent of how far these GM architectures truly learned the ground truth distribution. Furthermore, it has been debated whether striving for cutting-edge performance using these GMs provides appropriate research objectives to drive future algorithmic progress in generative artificial intelligence (GenAI) [Stein et al. \(2024\)](#).

Unlike discriminative models $P(T|X)$, which can be easily evaluated using a few labelled samples $(X_i; T_i)$, GM evaluation remains challenging. Although the likelihood function of a GM appears intuitive for assessment, it is generally intractable and scales poorly in higher dimensions. The GM evaluation literature abounds in metrics such as Inception score, Fréchet Inception distance, Kernel-Inception distance and a variety of k-nearest neighbour (KNN)-based two-value metrics, *precision* and *recall* (P&R) [Sajjadi et al. \(2018\)](#); [Simon et al. \(2019\)](#); [Kynkäänniemi et al. \(2019\)](#); [Naeem et al. \(2020\)](#); [Alaa et al. \(2022\)](#), improved P&R [Kynkäänniemi et al. \(2019\)](#), probabilistic [Park & Kim \(2023\)](#), and density and coverage [Naeem et al. \(2020\)](#) for GM evaluation. These metrics quantify the statistical divergence between the real $\{X_i\}$ and generated (fake) $\{Y_i\}$ samples using lower-dimensional representations. These are computed using a *Inception-V3* model trained on *ImageNet* dataset [Razavi et al. \(2019\)](#); [Karras et al. \(2020\)](#); [Dhariwal & Nichol \(2021\)](#); [Zheng et al. \(2021\)](#); [Anokhin et al. \(2021\)](#); [Chan et al. \(2021\)](#); [Ntavelis et al. \(2022\)](#). However, their over-reliance on *Inception-V3* embeddings makes them agnostic to features not related to the 1k classes of *ImageNet* [Heusel et al. \(2017\)](#); [Kynkäänniemi et al. \(2019\)](#); [Naeem et al. \(2020\)](#). In addition, these metrics are insensitive to fake distributional changes, unreliable due to outlier susceptibility, and rarely correlate with human perceptual scores [Zhou et al. \(2019\)](#); [Stein et al. \(2024\)](#). The absence of a

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

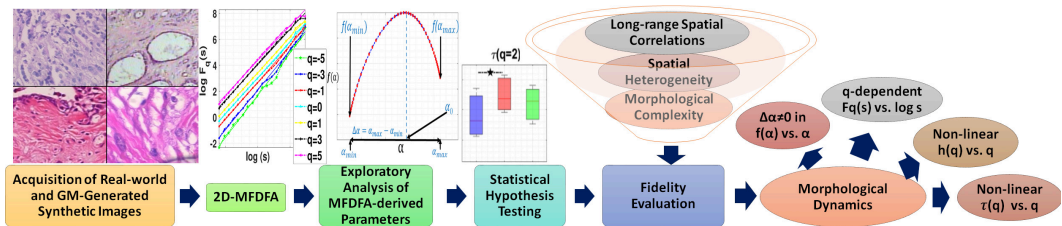


Figure 1: Schematic of multifractal analysis-based GM evaluation pipeline.

direct correlation with contextual morphological fidelity limits their practical applicability in critical scenarios such as medical imaging Aggarwal et al. (2021); Springenberg et al. (2023).

Medical imaging typically encodes multi-scale and non-linear micro-structural tissue convolutions with rich morphological diversity Das et al. (2014); Nawn et al. (2020). Similarly, industrial machine vision applications, such as hydrophobicity dielectrics or gas-insulated switchgear, include complex, multi-scale and non-linear spatial heterogeneity patterns Tang et al. (2015); Wang et al. (2025). Accordingly, GM-generated images must capture these task-specific fine-grained feature representations, which are arguably one of the most critical bottlenecks in GenAI applications Lopes & Betrouni (2009); Das et al. (2014); Nawn et al. (2020); Islam et al. (2013); Wu et al. (2023). Although prevalent GMs can produce photorealistic natural images, their efficacy in real-world medical imaging and industrial contexts remains challenging. In addition, standard GM evaluation metrics have yet to provide a clear consensus for these applications Yi et al. (2019); Wu et al. (2023). So, this work addresses the scientific question, 'To what extent do current GM architectures generate samples consistent with the morphological organization and complexity of real-world data?'

Beyond the spatial heterogeneity in GM-generated data, the origin of the *spectral bias* phenomenon in deep neural networks (DNNs) is not fully understood using the prevailing approaches, and still eludes researchers Fridovich-Keil et al. (2022). *spectral bias* is the propensity of DNNs to learn simple solutions for highly expressive functions that generalize surprisingly well. Consequently, DNNs are known to produce samples with elevated frequencies Rahaman et al. (2019); Schwarz et al. (2021). Thus, developing new techniques underpinning novel *structural* and *functional* aspects of a GM is essential for accelerating progress in generative AI Alaa et al. (2022). An active research area has been examining the dynamics of GM-generated data from a complexity science paradigm. These include image-space motion trajectories Li et al. (2024) and scaling laws in natural languages De Santis et al. (2023; 2024). It promises immense potential for applying sophisticated signal processing techniques for GM evaluation. Within this context, the concept of GM’s *multifractality* is introduced. Multifractal analysis is a *scaling* technique adapted from *statistical physics*, to characterize the power-law scaling dynamics. Although *power-law scaling* has been applied in medical imaging, primarily for representation learning Stosic & Stosic (2006); Lopes & Betrouni (2009); Islam et al. (2013); Das et al. (2014); Nawn et al. (2020), its potency for GM’s fidelity evaluation has not been explored to date and forms the primary motivation for this work.

Specifically, we develop a parametric evaluation framework for GM evaluation. Multiple statistically significant parameters are introduced in terms of multifractality and long-range spatial correlations. These parameters describe the morphological disparity between the synthesized and actual images. They also provide deeper insights into the complexity of generative structures. As intuited by non-linear dynamics, GM-generated images could be evaluated using the Hurst exponent and singularity spectrum width markers. To summarize, our main contributions are enumerated as:

- Proposing an information-theoretic framework (shown in Figure 1) to evaluate the fidelity of GM. The method uses multifractal analysis of GM-generated images. It proposes multiple metrics for the evaluation of GM in terms of complexity and long-range spatial correlation patterns to characterize their spatial dynamics according to the power law.
- Providing theoretical and numerical support for the evaluation metrics through experiments on real-world datasets with benchmark GM architectures. Our findings characterize the morphological organization and complexity of GM-generated images. It also provides deeper insights into *spectral bias* explainability.

2 PROPOSED METHODOLOGY AND PRELIMINARIES

Figure 1 outlines a detailed schematic of the proposed GM evaluation framework. The pipeline integrates the following modules: 2D-Multifractal Detrended Fluctuation Analysis (2D-MFDFA), extraction of multifractal markers from GM-generated synthesized images, their exploratory analysis, and hypothesis testing-based formulation of fidelity metrics. We briefly discuss the technical intricacies of different modules in the block diagram as follows.

2.1 2D-MFDFA

2D-MFDFA is a technique to examine the power-law scaling dynamics of a surface or image in terms of long-range spatial correlations and complexity of morphological structures [Kantelhardt et al. \(2002\)](#); [Xi et al. \(2016\)](#). An image is represented by a two-dimensional array, $X(i, j)$, where $i = 1, 2, \dots, N_1$ and $j = 1, 2, \dots, N_2$. 2D-MFDFA includes the following steps:

Step 1: $X(i, j)$, is divided into $N_{1s} \times N_{2s}$ disjointed square segments of size $s \times s$, where $N_{1s} = \lceil N_1/s \rceil$ and $N_{2s} = \lceil N_2/s \rceil$. These segments are represented as $X_{v,w} = X_{v,w}(i, j) = X(l_1 + i, l_2 + j) \forall 1 \leq i, j \leq s$, where $l_1 = (v-1)s$, $v = 1, 2, \dots, N_{1s}$, $l_2 = (w-1)s$, and $w = 1, 2, \dots, N_{2s}$.

Step 2: In each of these segments $X_{v,w}$, we compute a cumulative sum $u_{v,w}(i, j)$ as:

$$u_{v,w}(i, j) = \sum_{k_1=1}^i \sum_{k_2=1}^j X_{u,w}(k_1, k_2) \quad (1)$$

where $i, j = 1, 2, \dots, s$. Note that $u_{v,w}$ itself is a surface.

Step 3: The trend in the constructed surface, $u_{v,w}$ is computed by fitting a pre-determined bi-variate polynomial function \tilde{u} . Generally, a simple linear detrending function is used for detrending as:

$$\tilde{u}_{v,w}(i, j) = ai + bj + c, \quad (2)$$

where $1 \leq i, j \leq s$ and a, b and c are polynomial parameters to be estimated. Eqn. 2 denotes a plane characterized by simple matrix operations in the least-squares sense. Accordingly, the residual matrix, $\varepsilon_{i,j}$ is computed as:

$$\varepsilon_{i,j}(i, j) = u_{v,w}(i, j) - \tilde{u}_{v,w}(i, j) \quad (3)$$

The sample variance of the residual matrix $\varepsilon_{v,w}$ is used to compute the fluctuation function $F_{v,w}(s)$ for segment $X_{v,w}$

$$F_{v,w} = \sqrt{\frac{1}{s^2} \sum_{i=1}^s \sum_{j=1}^s (\varepsilon_{v,w}(i, j))^2} \quad (4)$$

Step 4: The overall detrended fluctuation function, $F_q(s)$ is computed by taking the mean of $F_{v,w}$ across all the segments and is calculated as:

$$F_q(s) = \left\{ \frac{1}{N_{1s}N_{2s}} \sum_{u=1}^{N_{1s}} \sum_{w=1}^{N_{2s}} F_{v,w}^q(s) \right\}^{\frac{1}{q}} \quad (5)$$

Where q can take any real value except $q = 0$. At $q = 0$, the factor $\frac{1}{q}$ in Eqn. 5 blows up, so normal averaging fails. Instead, a logarithmic averaging procedure using L'Hôspital's rule is applied to compute $F_0(s)$, and is calculated as:

$$F_0(s) = \exp \left\{ \frac{1}{N_{1s}N_{2s}} \sum_{u=1}^{N_{1s}} \sum_{w=1}^{N_{2s}} \ln[F_{u,w}(s)] \right\} \quad (6)$$

Step 5: We vary the scale size, s in $[s_{min}, s_{max}]$ where $s_{min} = 10$ to $s_{max} = \min[N_1, N_2]/4$ and compute the power-law scaling relationship between $F_q(s)$ and s ,

$$F_q(s) \sim s^{h(q)} \quad (7)$$

A surface exhibiting power-law scaling behaviour will have $\log F_q$ linearly depending on $\log s$, with $h(q)$ as the slope. Here, $h(q)$ is the Generalized Hurst Exponent. $h(q)$ characterizes the scaling nature of the segments with large (small) fluctuations for positive (negative) exponent values of q . The value of $h(q = 2)$ is known as the Hurst Exponent, which characterizes a surface’s long-range persistence or memory. It is related to the decay rate of the statistical dependence between two points on a surface with an increasing gap between them [Ashkenazy et al. \(2002\)](#). Moreover, $h(q)$ is related to the classical scaling exponent, $\tau(q)$ as:

$$\tau(q) = qh(q) - D_f \quad (8)$$

where D_f is the fractal dimension of the geometric support of the multifractal measure and $D_f = 2$ for a surface.

$h(q)$ is independent of q and $\tau(q)$ is linearly dependent on q for a monofractal surface. So, a unique value of $h(q)$ characterizes the spatial dynamics of a monofractal surface [Kantelhardt et al. \(2002\)](#). In contrast, $h(q)$ is a function of q , and $\tau(q)$ varies non-linearly with q for a multifractal surface [Ashkenazy et al. \(2003\)](#). Moreover, the singularity spectrum, $f(\alpha)$ for a multifractal surface is related to $\tau(q)$ by Legendre transform [Shao & Ditlevsen \(2016\)](#), and is given by:

$$\alpha = d\tau(q)/dq, \quad f(\alpha) = q\alpha - \tau(q) \quad (9)$$

Where α is the Hölder exponent and $f(\alpha)$ is the multifractal spectrum. $f(\alpha)$ characterizes the relative relative strength of the range of fractal exponents (α) to the analyzed surface [Shimizu et al. \(2002\)](#). Using Eqn. 8, we can write Eqn. 9 as:

$$\alpha = h(q) + qh'(q), \quad f(\alpha) = q[\alpha - h(q)] + 2 \quad (10)$$

Usually, a quadratic function is fitted to the multifractal spectra around the position of maximum α_0 for analysis as:

$$f(\alpha) = A(\alpha - \alpha_0)^2 + B(\alpha - \alpha_0) + C \quad (11)$$

Where C is an additive constant, whose value is given by $C = f(\alpha_0)$. Here, B signifies the singularity spectrum asymmetry and is zero for a symmetric spectrum, while A indicates the concavity ($A < 0$) or convexity ($A > 0$) of the singularity spectrum. The singularity spectrum width, $\Delta\alpha$ of the singularity spectrum is computed by extrapolating Eqn. 11 to zero. $\Delta\alpha$ is computed as:

$$\Delta\alpha = \alpha_{max} - \alpha_{min}, \quad s.t. f(\alpha_{max}) = f(\alpha_{min}) = 0 \quad (12)$$

Here, α_{max} and α_{min} correspond to the right and left extremities of $f(\alpha)$ on the $f(\alpha)$ vs. α plot. The value of $\Delta\alpha$ characterizes the degree of multifractality of an image and correlates with its singularity strength [Shimizu et al. \(2002\)](#). For a monofractal surface, $h(q)$ is independent of q , which is equivalent to a unique value of $f(\alpha)$ describing the multifractal spectrum $\forall\alpha$; hence, $\Delta\alpha = 0$ (following Eqn. 10 and 12).

2.2 MULTIFRACTAL MARKERS FOR FIDELITY EVALUATION

Figure 2 demonstrates a typical $f(\alpha)$ vs. α plot exemplifying the multifractal spectrum and different shape parameters derived from its geometry. These parameters characterize the morphological organization of an image in terms of the complexity of morphological structures [Tang et al. \(2015\)](#); [Nawn et al. \(2020\)](#). The significance of different parameters extracted from 2D-MFDFA for fidelity evaluation is described in Table 1.

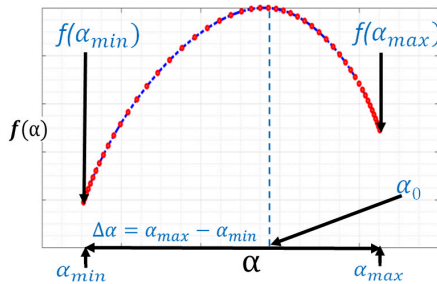


Figure 2: Illustration of different shape parameters extracted from a typical multifractal spectrum, i.e., plot of $f(\alpha)$ vs. α .

Marker	Description
$h(q = 2)$	Indicate the nature of long-range spatial correlations.
α_{min}	Right extremity of the singularity exponent.
α_{max}	Left extremity of the singularity exponent.
$f(\alpha_{min})$	Singularity spectrum value at α_{min} .
$f(\alpha_{max})$	Singularity spectrum value at α_{max} .
$\Delta\alpha$	Singularity strength. Indicate the degree of multifractality.
α_0	Location of Singularity spectrum’s maxima/minima.
$f(\alpha_0)$	Singularity spectrum value at α_0 .
A	Singularity spectrum’s concavity ($A < 0$) or convexity ($A > 0$).
B	Singularity spectrum asymmetry.

Table 1: Description of extracted multifractal markers.

All the multifractal markers described in Table 1 are used to evaluate the fidelity of the GM-generated images. Here, $h(q = 2)$ characterizes the long-range spatial correlations in an image. α_{min} and α_{max} exemplify the bounds of the singularity exponent, α in the 2D-MFDFA investigation Nawn et al. (2020). $f(\alpha_{min})$ and $f(\alpha_{max})$ indicate the singularity spectrum value at the left and right extremities of the singularity exponent α in $f(\alpha)$ vs. α plot. $\Delta\alpha$ indicates an image’s singularity strength computed between the extremities of α . $\Delta\alpha = 0$ for a monofractal surface, and larger values correlate with a higher degree of multifractality. A higher value of α_0 signifies an image’s highly coherent and regular morphological structures. $f(\alpha_0)$ indicates the maximum strength of the singularity exponent and is a measure of roughness in an image. The singularity spectrum asymmetry index, B (Eqn. 11) describes the asymmetry of the multifractal spectrum geometry. A positive B is associated with lower-order fractal exponents (i.e., the dominance of positive q values), whereas a negative B indicates the dominance of negative q values. Except for $h(q = 2)$ in Table 1, the multifractal spectrum geometry parameters assess an image’s multifaceted aspect of morphological organization. The fidelity of the GM-generated synthesized images is evaluated in terms of derived multifractal markers’ disparity from real-world images. Moreover, to relax the oversimplified assumptions, we use hypothesis testing-based assessment of the multifractal markers in a probabilistic sense.

2.3 EXPLORATORY FEATURE ANALYSIS AND HYPOTHESIS TESTS

We use BoxPlot representation of the derived multifractal markers for visualization of summary statistics (locality, spread, and skewness) Frigge et al. (1989) and use non-parametric statistics-based hypothesis tests Fagerland (2012) for inferential statistical analysis. Non-parametric hypothesis tests have been used to avoid making assumptions about the underlying data distribution. In particular, we use the Wilcoxon rank sum test Woolson (2005) to compare the location (median; η) of the multifractal markers across the study groups (G_i) for different datasets and compute the statistical significance of the null hypothesis: $\eta_{G_i} = \eta_{G_j}$, against the alternative hypothesis: $(\eta_{G_i} < \eta_{G_j}) \vee (\eta_{G_i} > \eta_{G_j})$ at a standard significance level of 0.5% Hosmer & Lemeshow (1992). Here, G_i corresponds to the various GM architectures explored in the experimental evaluation. We demonstrate the sensitivity of statistically significant evaluation metrics on the BoxPlots by using $- * -$.

3 EXPERIMENTAL EVALUATION

Amidst various imaging modalities, multi-scale morphological structures are noteworthy in the microscopic domain Kaji et al. (2024); Qian et al. (2012); Braumann et al. (2005); Das et al. (2014); Nawn et al. (2020). So, we verified the efficacy of our proposed GM evaluation approach on three real-world challenging histopathological datasets (*BreakHis*, *BACH*, and *Oral Cancer*) and across benchmark GM architecture, namely, *GAN*, *VAE*, *NF*, and *Diffusion*.

Datasets. *BreakHis* contain 9,109 (2,480 benign and 5,429 malignant) microscopic images of breast tumour tissues collected from 82 patients at magnifying factors of 40X, 100X, 200X, and 400X Spanhol et al. (2015). *BACH* include 400+ whole-slide microscopy images from four classes, namely, normal, benign, in situ carcinoma and invasive carcinoma Aresta et al. (2019); Araújo et al. (2017); Fondón et al. (2018). *Oral Cancer* comprise 1224 (290 normal epithelium and 934 oral squamous cell carcinoma) histopathological images at 100X and 400X magnification levels Rahman et al. (2020).

Baseline GMs. We have used a deep convolutional GAN Radford et al. (2016) to generate the GAN network’s images. The generator in the GAN architecture consisted of five upsampling blocks with transposed convolutions, while the discriminator had four downsampling blocks with normal convolutions. Similarly, we used the VAE model in Razavi et al. (2019) to generate the VAE network’s images. In our experiments with NF, we used a Flow++ Ho et al. (2019) model with two dequantization blocks and ten residual blocks and used checkerboard coupling layers. The NF model was trained using an Adam Optimizer with a negative log-likelihood loss function. We used a denoising diffusion probabilistic model (DDPM) Ho et al. (2020a) to generate the diffusion-based GM images. 2D UNet from huggingface diffusers¹ has been used as the underlying generator network consisting of six downsampling blocks and restrained upsampling blocks in the DDPM architecture,

¹<https://huggingface.co/>

with the downsampling and upsampling blocks includes a spatial self-attention mechanism. The DDPM was trained using an Adam Optimizer and mean square error loss function.

Implementation Details. An image resolution of 128×128 was used for training and inference across all the three datasets. Since histopathology images encode intricate cellular architecture of tissues, standard data augmentation through rotation or shearing may distort the anatomical structures, leading to unrealistic variations and histological inconsistency Wei et al. (2019). So, only random horizontal and vertical flips were used to augment the dataset size. No pre-training was performed on the generator model used to generate the images. All the experiments were done on a single 80 GB Nvidia A100 GPU with FP16 precision, and all the GMs were trained with a maximum number of training epochs set to 500. For all the experiments, we have used AdamW optimizer ($\alpha = 0.0001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$) with a weight decay factor of 0.01. A batch size of 128 and 512 were used during training and data generation, respectively. We generate 2000 images for each dataset with resolution 128×128 , the same as the input images.

3.1 EVIDENCE OF POWER-LAW SCALING DYNAMICS.

Figure 3 illustrates the typical variation of $\log F_q(s)$ vs. $\log s$ for different moments (q), obtained from 2D-MFDFA on *Original* images taken from Breakhis (Figure 3a), BACH (Figure 3b), and Oral Cancer (Figure 3c) datasets, respectively and the corresponding synthetic images generated by different GM architectures. $\log F_q$ linearly depends on $\log s$, with $h(q)$ as the slope, which signifies the power-law scaling behaviour. Moreover, these scaling patterns are consistently observed across all the *Original* images and those generated by different GM architectures, thereby suggesting further analysis of the generalized Hurst exponent, $h(q)$ (Eqn. 7) and singularity spectrum, $f(\alpha)$ vs. α (Eqn. 12) for characterizing the GM-generated image dynamics.

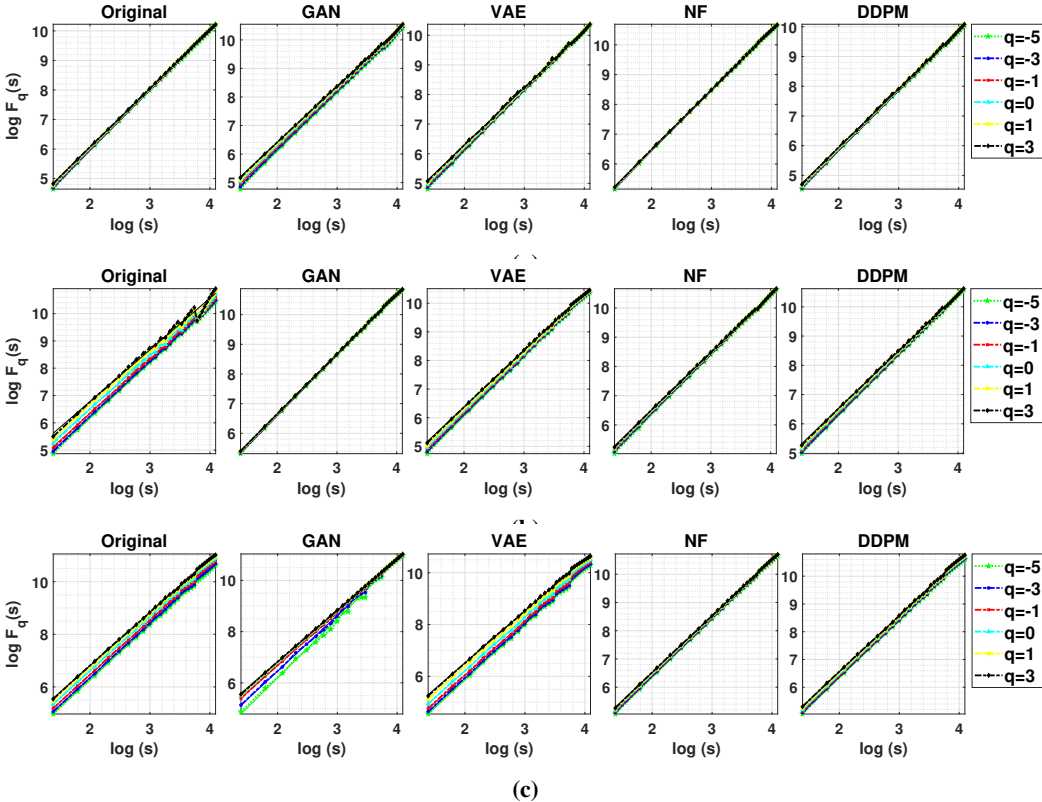


Figure 3: Typical variation of $\log F_q(s)$ vs. $\log s$ for different q , obtained using 2d-MFDFA of the *Original* image samples and synthetic images generated by different GM architectures. The GMs are trained on (a) Breakhis, (b) BACH, and (c) Oral Cancer datasets, respectively. Here, $\log F_q$ linearly depending on $\log s$, with $h(q)$ as the slope, indicates power-law scaling behaviour.

The typical variation of $\log F_q(s)$ vs. $\log s$ for different moments (q), obtained in Figure 3a for the Breakhis dataset is uniform across all the *Original* images and those generated by different GM architectures, with GAN showing slight (q)-dependent slope variations. Similarly, the *Original* images from the BACH dataset exhibits (q)-dependent slope variations in the $\log F_q(s)$ vs. $\log s$ plots, which is comparatively less discernible in different GM-generated images in Figure 3b. Lastly, for the Oral Cancer dataset in Figure 3c, the *Original* image exhibits (q)-dependent slope variations in the $\log F_q(s)$ vs. $\log s$ plots, which is reciprocated in different GM-generated images.

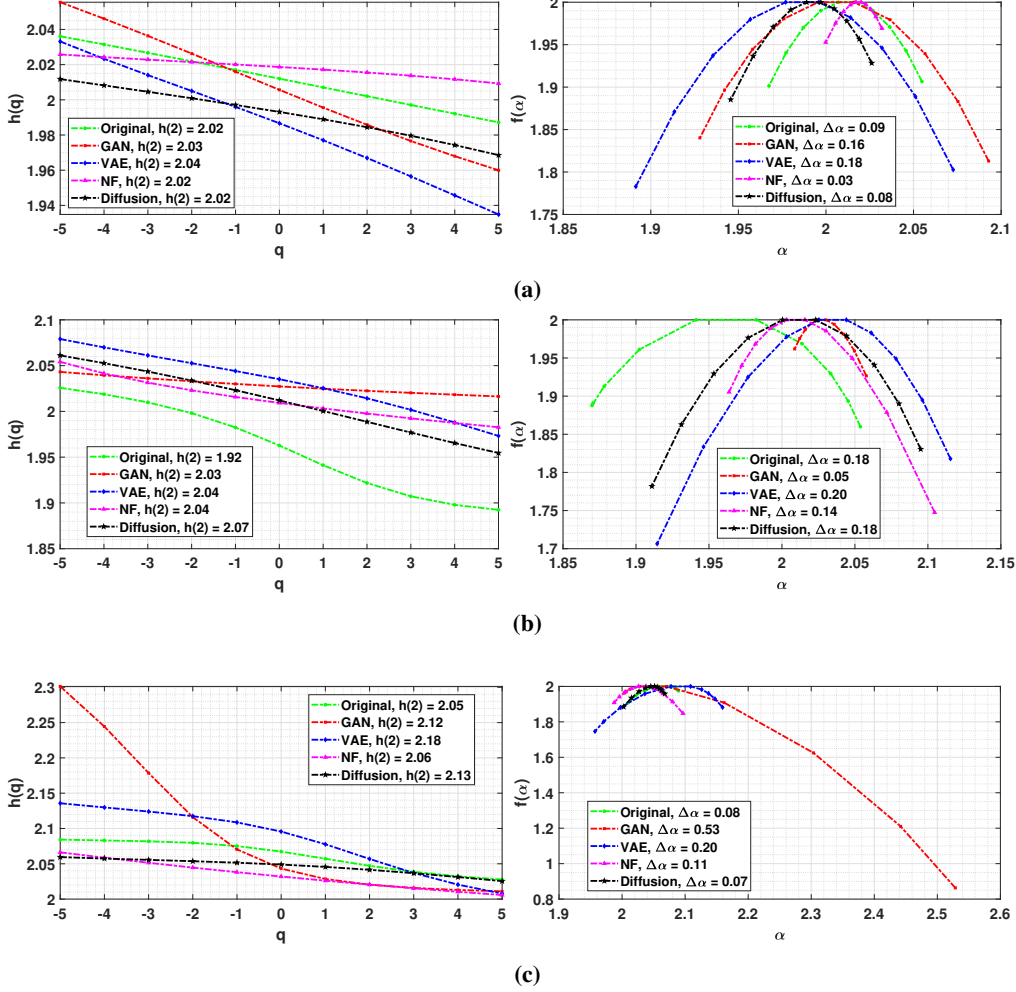


Figure 4: Typical (q)-dependent generalized Hurst exponent $h(q)$ and singularity spectrum $f(\alpha)$ vs. α is shown for *Original* and GM-generated samples generated from 2D-MFDFA of images, taken from (a) Breakhis, (b) BACH, and (c) Oral Cancer datasets, respectively. Here, the multifractality in the *Original* and GM-generated synthesized images is evident from non-linear $h(q)$ vs. q and $\Delta\alpha \neq 0$.

3.2 EVIDENCE OF MULTIFRACTALITY FROM $h(q)$ VS. q AND $f(\alpha)$ VS. α ANALYSIS

Figure 4 demonstrate the typical q -dependent $h(q)$ and $f(\alpha)$ vs. α variations, obtained from 2D-MFDFA on *Original* images taken from Breakhis (Figure 4a), BACH (Figure 4b), and Oral Cancer (Figure 4c) datasets, respectively and the corresponding synthetic images generated by different GM architectures. Our results highlight decreasing $h(q)$ with increasing q and $h(q)$ varying non-linearly for negative and positive orders of q . This non-linear variation of $h(q)$ with q and $\Delta\alpha \neq 0$ (Refer to Eqn. 12) patterns are consistently observed across all the *Original* images and synthesized images

generated by different GM architectures, thereby highlighting their underlying multifractal nature. Also, the typical values of $h(q = 2) \approx 2$ suggest the presence of long-range spatial correlation patterns, which are persistent for both the *Original* and synthetic images generated by different GM architectures. $\Delta\alpha \neq 0$ from $f(\alpha)$ vs. α variations suggest the analysis of different sources of morphological complexity in an image Zhou et al. (2008).

Figure 5 demonstrates the BoxPlot-based visualization of the derived multifractal markers, $h(q = 2)$, α_{min} , α_{max} , $f(\alpha_{min})$, $f(\alpha_{max})$, $\Delta\alpha$, and the singularity spectrum asymmetry index, B (Eqn. 11), obtained from 2D-MFDFA of *Original* images taken from the Oral Cancer datasets and synthesized images. Across different GM architectures, our study shows that DDPM-based diffusion models come in close agreement with the *Original* images in terms of the underlying complexity of the morphological structures. In the next section, we study the different sources of complexity and quantify the fidelity in terms of different multifractal markers.

3.3 SOURCES OF MORPHOLOGICAL COMPLEXITY

Figure 6 demonstrate the BoxPlot-based visualization of and different sources of complexity from $\Delta\alpha$, $\Delta\alpha_{Shuffled}$, and $\Delta\alpha_{Density}$, obtained from 2D-MFDFA of GM-generated (Diffusion & GAN) and real-world (Original) images taken from Breakhis (1st row), BACH (2nd row), and Oral Cancer (3rd row) datasets, respectively. Here, $\Delta\alpha_{Density}$ is the singularity spectrum width after randomly shuffling all the data points in an image surface and analyzing the multifractality. Shuffling destroys all the long-range correlations, and the resulting source of multifractality is due to the data distribution. $\Delta\alpha_{Shuffled}$ is the difference between the singularity spectrum width obtained from the shuffled and original image. We argue that the spectral bias phenomenon in DNNs is due to the data distribution

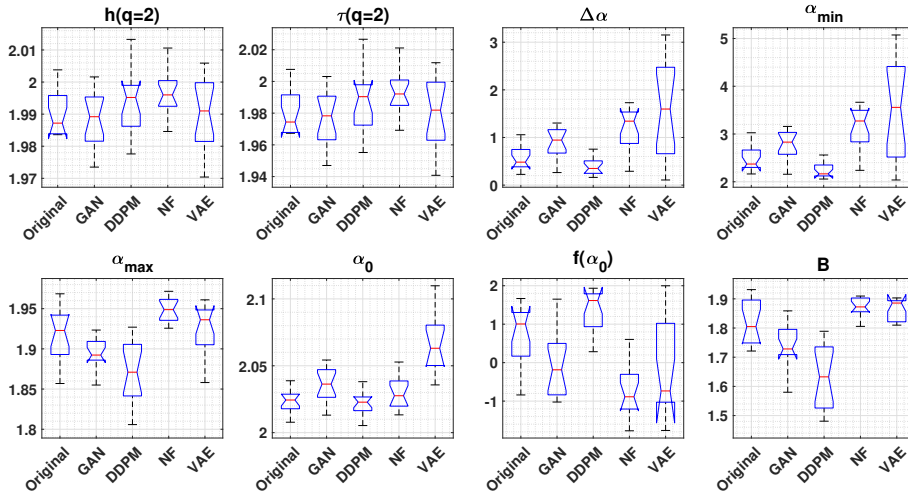


Figure 5: BoxPlots showing the variation (locality, spread, and skewness) of the derived multifractal markers, $h(q = 2)$, α_{min} , α_{max} , $f(\alpha_{min})$, $f(\alpha_{max})$, $\Delta\alpha$, and singularity spectrum asymmetry index, B (Eqn. 11), obtained from 2D-MFDFA of *Original* images taken from the Oral Cancer datasets and synthesized images using different GM architectures.

4 CONCLUSION

This work proposed a framework for systematically assessing the fidelity of GM-generated images through the lens of multifractality. Multiple multifractal markers are introduced using non-parametric hypothesis testing for GM evaluation. These metrics characterize long-range correlated dynamics and structural complexity in GM-generated images. They quantify the statistical disparity in the morphological organization between the synthesized and actual data. Experimental results reveal consistent multifractal patterns across GM architectures and real-world datasets. Our analysis provides deeper insight into the origins of the underlying complexity of GM and offers a plausible explanation

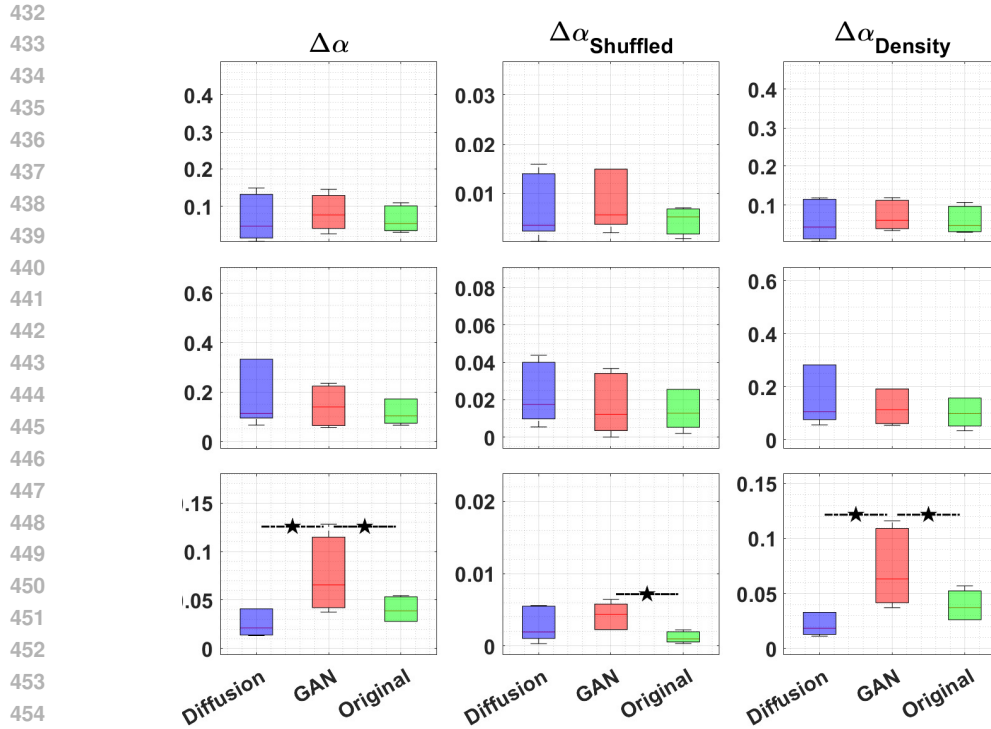


Figure 6: BoxPlots showing the variation (locality, spread, and skewness) of the derived multifractal markers ($h(q=2)$, $\Delta\alpha$, $\Delta\alpha_{Shuffled}$, and $\Delta\alpha_{Density}$), obtained from 2D-MFDFA of GM-generated (Diffusion & GAN) and real-world (Original) HI for Breakhis, BACH, and Oral Cancer datasets, respectively. The inter-class statistical difference using the non-parametric Wilcoxon rank sum test at a 0.5% significance level is shown with $- * -$.

for the spectral bias phenomenon. We believe that the proposed metrics will primarily benefit medical imaging research. They enable the fairness assessment of GM-generated histopathological images, where multi-scale deformation dynamics of multifarious micro-structural tissue convolutions are encoded due to long-range spatial correlations. These patterns alter during pathological processes in patient subgroups. We leave an in-depth study of this use case for future work. We also plan to use information-theoretic measures, such as recurrence quantification, for GM evaluation. In addition, our multifractal formalism can easily be extended to 1D and high-dimensional settings.

REFERENCES

- Ravi Aggarwal, Viknesh Sounderajah, Guy Martin, Daniel SW Ting, Alan Karthikesalingam, Dominic King, Hutan Ashrafian, and Ara Darzi. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ digital medicine*, 4(1):65, 2021.
- Ahmed Alaa, Boris Van Breugel, Evgeny S Saveliev, and Mihaela van der Schaar. How faithful is your synthetic data? sample-level metrics for evaluating and auditing generative models. pp. 290–306, 2022.
- Ivan Anokhin, Kirill Demochkin, Taras Khakhulin, Gleb Sterkin, Victor Lempitsky, and Denis Korzhenkov. Image generators with conditionally-independent pixel synthesis. pp. 14278–14287, 2021.
- Teresa Araújo, Guilherme Aresta, Eduardo Castro, José Rouco, Paulo Aguiar, Catarina Eloy, António Polónia, and Aurélio Campilho. Classification of breast cancer histology images using convolutional neural networks. *PLoS one*, 12(6):e0177544, 2017.

- 486 Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan,
487 Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al. Bach:
488 Grand challenge on breast cancer histology images. *Medical image analysis*, 56:122–139, 2019.
489
- 490 Yosef Ashkenazy, Jeffrey M Hausdorff, Plamen Ch Ivanov, and H Eugene Stanley. A stochastic
491 model of human gait dynamics. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):
492 662–670, 2002.
- 493 Yosef Ashkenazy, Shlomo Havlin, Plamen Ch Ivanov, Chung-K Peng, Verena Schulte-Frohlinde, and
494 H Eugene Stanley. Magnitude and Sign Scaling in Power-Law Correlated Time Series. *Physica A:
495 Statistical Mechanics and its Applications*, 323:19–41, 2003.
496
- 497 U-D Braumann, J-P Kuska, Jens Eickenel, L-C Horn, Markus Loffler, and Michael Hockel. Three-
498 dimensional reconstruction and quantification of cervical carcinoma invasion fronts from histologi-
499 cal serial sections. *IEEE Transactions on Medical Imaging*, 24(10):1286–1307, 2005.
- 500 Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. pi-gan: Periodic
501 implicit generative adversarial networks for 3d-aware image synthesis. pp. 5799–5809, 2021.
502
- 503 Nandan Das, Subhasri Chatterjee, Satish Kumar, Asima Pradhan, Prasanta Panigrahi, I Alex Vitkin,
504 and Nirmalya Ghosh. Tissue Multifractality and Born Approximation in Analysis of Light
505 Scattering: a Novel Approach for Precancers Detection. *Scientific reports*, 4(1):1–7, 2014.
506
- 507 Enrico De Santis, Giovanni De Santis, and Antonello Rizzi. Multifractal characterization of texts
508 for pattern recognition: on the complexity of morphological structures in modern and ancient
509 languages. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10143–10160,
510 2023.
- 511 Enrico De Santis, Alessio Martino, and Antonello Rizzi. Human versus machine intelligence:
512 Assessing natural language generation models through complex systems theory. *IEEE Transactions
513 on Pattern Analysis and Machine Intelligence*, 2024.
- 514 Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances
515 in neural information processing systems*, 34:8780–8794, 2021.
516
- 517 Morten W Fagerland. t-tests, Non-parametric Tests, and Large Studies—A Paradox of Statistical
518 Practice? *BMC Medical Research Methodology*, 12:1–7, 2012.
- 519 Irene Fondón, Auxiliadora Sarmiento, Ana Isabel García, María Silvestre, Catarina Eloy, António
520 Polónia, and Paulo Aguiar. Automatic classification of tissue malignancy for breast carcinoma
521 diagnosis. *Computers in biology and medicine*, 96:41–51, 2018.
522
- 523 Sara Fridovich-Keil, Raphael Gontijo Lopes, and Rebecca Roelofs. Spectral bias in practice: The
524 role of function frequency in generalization. *Advances in Neural Information Processing Systems*,
525 35:7368–7382, 2022.
- 526 Michael Frigge, David C Hoaglin, and Boris Iglewicz. Some Implementations of the Boxplot. *The
527 American Statistician*, 43(1):50–54, 1989.
528
- 529 Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair,
530 Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the
531 ACM*, 63(11):139–144, 2020.
532
- 533 Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans
534 trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural
535 information processing systems*, 30, 2017.
- 536 Jonathan Ho, Xi Chen, Aravind Srinivas, Yan Duan, and Pieter Abbeel. Flow++: Improving flow-
537 based generative models with variational dequantization and architecture design. 97:2722–2730,
538 09–15 Jun 2019. URL <https://proceedings.mlr.press/v97/ho19a.html>.
539
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. 2020a.

- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020b.
- David W Hosmer and Stanley Lemeshow. Confidence Interval Estimation of Interaction. *Epidemiology*, pp. 452–456, 1992.
- Atiq Islam, Syed MS Reza, and Khan M Iftekharuddin. Multifractal Texture Estimation for Detection and Segmentation of Brain Tumors. *IEEE transactions on biomedical engineering*, 60(11):3204–3215, 2013.
- Shizuo Kaji, Naoya Tanabe, Tomoki Maetani, Yusuke Shiraishi, Ryo Sakamoto, Tsuyoshi Oguma, Katsuhiko Suzuki, Kunihiko Terada, Motonari Fukui, Shigeo Muro, et al. Quantification of airway structures by persistent homology. *IEEE Transactions on Medical Imaging*, 2024.
- Jan W Kantelhardt, Stephan A Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H Eugene Stanley. Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.
- Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. pp. 4401–4410, 2019.
- Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33:12104–12114, 2020.
- Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34:852–863, 2021.
- Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018.
- Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems*, 32, 2019.
- Zhengqi Li, Richard Tucker, Noah Snively, and Aleksander Holynski. Generative image dynamics. pp. 24142–24153, 2024.
- Renaud Lopes and Nacim Betrouni. Fractal and Multifractal Analysis: a Review. *Medical image analysis*, 13(4):634–649, 2009.
- Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. pp. 7176–7185, 2020.
- Debaleena Nawn, Sawon Pratiher, Subhankar Chattoraj, Debjani Chakraborty, Mousumi Pal, Ranjan Rashmi Paul, Srimonti Dutta, and Jyotirmoy Chatterjee. Multifractal Alterations in Oral Sub-Epithelial Connective Tissue During Progression of Pre-Cancer and Cancer. *IEEE Journal of Biomedical and Health Informatics*, 25(1):152–162, 2020.
- Evangolos Ntavelis, Mohamad Shahbazi, Iason Kastanis, Radu Timofte, Martin Danelljan, and Luc Van Gool. Arbitrary-scale image synthesis. pp. 11533–11542, 2022.
- Dogyun Park and Suhyun Kim. Probabilistic precision and recall towards reliable evaluation of generative models. pp. 20099–20109, 2023.
- Ai-Rong Qian, D Li, J Han, X Gao, S Mi Di, W Zhang, LF Hu, and Peng Shang. Fractal dimension as a measure of altered actin cytoskeleton in mc3t3-e1 cells under simulated microgravity using 3-d/2-d clinostats. *IEEE Transactions on Biomedical Engineering*, 59(5):1374–1380, 2012.

- 594 Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep
595 convolutional generative adversarial networks, 2016. URL [https://arxiv.org/abs/1511.](https://arxiv.org/abs/1511.06434)
596 [06434](https://arxiv.org/abs/1511.06434).
- 597 Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua
598 Bengio, and Aaron Courville. On the spectral bias of neural networks. pp. 5301–5310, 2019.
- 600 Tabassum Yesmin Rahman, Lipi B Mahanta, Anup K Das, and Jagannath D Sarma. Histopathological
601 imaging database for oral cancer analysis. *Data in brief*, 29:105114, 2020.
- 602 Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with
603 vq-vae-2. *Advances in neural information processing systems*, 32, 2019.
- 604 Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-
605 resolution image synthesis with latent diffusion models. pp. 10684–10695, 2022.
- 606 Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing
607 generative models via precision and recall. *Advances in neural information processing systems*, 31,
608 2018.
- 611 Katja Schwarz, Yiyi Liao, and Andreas Geiger. On the frequency bias of generative models. *Advances*
612 *in Neural Information Processing Systems*, 34:18126–18136, 2021.
- 613 Zhi-Gang Shao and Peter D Ditlevsen. Contrasting scaling properties of interglacial and glacial
614 climates. *Nature Communications*, 7(1):10951, 2016.
- 615 YU Shimizu, Stefan Thurner, and Klaus Ehrenberger. Multifractal Spectra as a Measure of Complexity
616 in Human Posture. *Fractals*, 10(01):103–116, 2002.
- 617 Loic Simon, Ryan Webster, and Julien Rabin. Revisiting precision recall definition for generative
618 modeling. pp. 5799–5808, 2019.
- 619 Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben
620 Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint*
621 *arXiv:2011.13456*, 2020.
- 622 Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of
623 score-based diffusion models. *Advances in Neural Information Processing Systems*, 34:1415–1428,
624 2021.
- 625 Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast
626 cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):
627 1455–1462, 2015.
- 628 Maximilian Springenberg, Annika Frommholz, Markus Wenzel, Eva Weicken, Jackie Ma, and Nils
629 Strodthoff. From modern cnns to vision transformers: Assessing the performance, robustness, and
630 classification strategies of deep learning models in histopathology. *Medical Image Analysis*, 87:
631 102809, 2023.
- 632 George Stein, Jesse Cresswell, Rasa Hosseinzadeh, Yi Sui, Brendan Ross, Valentin Vilecroze,
633 Zhaoyan Liu, Anthony L Caterini, Eric Taylor, and Gabriel Loaiza-Ganem. Exposing flaws of
634 generative model evaluation metrics and their unfair treatment of diffusion models. *Advances in*
635 *Neural Information Processing Systems*, 36, 2024.
- 636 Tatijana Stosic and Boriko D Stosic. Multifractal Analysis of Human Retinal Vessels. *IEEE*
637 *transactions on medical imaging*, 25(8):1101–1107, 2006.
- 638 Ju Tang, Dibo Wang, Lei Fan, Ran Zhuo, and Xiaoxing Zhang. Feature parameters extraction of gis
639 partial discharge signal with multifractal detrended fluctuation analysis. *IEEE Transactions on*
640 *Dielectrics and Electrical Insulation*, 22(5):3037–3045, 2015.
- 641 Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space.
642 *Advances in Neural Information Processing Systems*, 34:11287–11302, 2021.

648 Yiwei Wang, Li Zhang, Bilal Iqbal Ayubi, and Guowei Hou. Experimental methods for power fre-
649 quencyhigh frequency equivalency based on dielectric strength and partial discharge characteristics
650 of insulation materials. *IEEE Transactions on Dielectrics and Electrical Insulation*, 2025.

651 Jerry Wei, Arief Suriawinata, Louis Vaickus, Bing Ren, Xiaoying Liu, Jason Wei, and Saeed
652 Hassanpour. Generative image translation for data augmentation in colorectal histopathology
653 images. *Proceedings of machine learning research*, 116:10, 2019.

654 Robert F Woolson. Wilcoxon signed-rank test. *Encyclopedia of Biostatistics*, 8, 2005.

655 Weiwen Wu, Yanyang Wang, Qiegen Liu, Ge Wang, and Jianjia Zhang. Wavelet-improved score-
656 based generative model for medical imaging. *IEEE Transactions on Medical Imaging*, 2023.

657 Caiping Xi, Shuning Zhang, Gang Xiong, and Huichang Zhao. A comparative study of two-
658 dimensional multifractal detrended fluctuation analysis and two-dimensional multifractal detrended
659 moving average algorithm to estimate the multifractal spectrum. *Physica A: Statistical Mechanics
660 and its Applications*, 454:34–50, 2016.

661 Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review.
662 *Medical image analysis*, 58:101552, 2019.

663 Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. The spatially-correlative loss for various image
664 translation tasks. pp. 16407–16417, 2021.

665 Sharon Zhou, Mitchell Gordon, Ranjay Krishna, Austin Narcomey, Li F Fei-Fei, and Michael
666 Bernstein. Hype: A benchmark for human eye perceptual evaluation of generative models.
667 *Advances in neural information processing systems*, 32, 2019.

668 Wei-Xing Zhou et al. Multifractal Detrended Cross-Correlation Analysis for Two Nonstationary
669 Signals. *Physical Review E*, 77(6):066211, 2008.

670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701