An Interpretable Evaluation of Entropy-based Novelty of Generative Models

Jingwei Zhang [1] **Cheuk Ting Li** [2] **Farzan Farnia** [1]

Abstract

The massive developments of generative model frameworks require principled methods for the evaluation of a model's novelty compared to a reference dataset. While the literature has extensively studied the evaluation of the quality, diversity, and generalizability of generative models, the assessment of a model's novelty compared to a reference model has not been adequately explored in the machine learning community. In this work, we focus on the novelty assessment for multi-modal distributions and attempt to address the following *differential clustering* task: Given samples of a generative model $P_{\mathcal{G}}$ and a reference model P_{ref}, how can we discover the sample types expressed by $P_{\mathcal{G}}$ more frequently than in P_{ref}? We introduce a spectral approach to the differential clustering task and propose the *Kernel-based Entropic Novelty (KEN)* score to quantify the mode-based novelty of $P_{\mathcal{G}}$ with respect to P_{ref}. We analyze the KEN score for mixture distributions with well-separable components and develop a kernel-based method to compute the KEN score from empirical data. We support the KEN framework by presenting numerical results on synthetic and real image datasets, indicating the framework's effectiveness in detecting novel modes and comparing generative models. The paper's code is available at: `github.com/buyeah1109/KEN`.

1. Introduction

Deep generative models including variational autoencoders (VAEs) (Kingma & Welling, 2013), generative adversarial networks (GANs) (Goodfellow et al., 2014), and denoising diffusion models (Ho et al., 2020) have attained remarkable results in many machine learning problems. The success of these models is primarily due to the great capacity of deep neural networks to express the complex distributions of image, audio, and text data. The impressive qualitative results of deep generative models have inspired several theoretical and empirical studies on their evaluation to reveal the advantages and disadvantages of existing architectures for training generative models.

To compare different generative modeling schemes, several evaluation metrics have been proposed in the literature. The existing evaluation scores can be classified into two categories: 1) distance-based metrics such as Fréchet Inception Distance (FID) (Heusel et al., 2017) and Kernel Inception Distance (KID) (Bińkowski et al., 2018) measuring the closeness of the distribution of data and generative model, 2) quality, diversity, and generalizability scores such as Inception score (Salimans et al., 2016), Precision/Recall (Sajjadi et al., 2018; Kynkäänniemi et al., 2019), and Density/Coverage (Naeem et al., 2020) assessing the sharpness and variety of the generated data. The mentioned metrics tend to assign higher scores to models closer to the underlying data distribution. While such a property is desired in the evaluation of a learning framework, it may not result in an assessment of a generative model's novelty compared to a baseline generative model or another reference distribution.

However, the massive developments of generative models highlight the need to assess a model's novelty compared to other models, because an interpretable comparison between generative models requires the identification of sample types generated by one model more frequently than by the other models. Moreover, prompt-based generative models are often utilized to follow the user's input text prompts to create novel contents, e.g. images of a novel scene or object. If the goal is to maximize the uncommonness of the generated data compared to a reference dataset, a relevant evaluation factor is the model's expressed novelty in comparison to the reference distribution.

In this work, we focus on the novelty evaluation task in the context of multi-modal distributions which are often present in large-scale image and text datasets due to the different background features of the collected data. In our theoretical analysis, we suppose the test and reference models consist

[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong, [2]Department of Information Engineering, The Chinese University of Hong Kong. Correspondence to: Jingwei Zhang <jwzhang22@cse.cuhk.edu.hk>, Cheuk Ting Li <ctli@ie.cuhk.edu.hk>, Farzan Farnia <farnia@cse.cuhk.edu.hk>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

of multiple modes and aim to solve a *differential clustering* task for identifying the novel modes produced by the test model more frequently than by the reference distribution. We propose a *spectral approach* to the differential clustering problem by analyzing the *kernel covariance matrix* of the test and reference distributions, yielding eigenvalues measuring the frequency of differently-expressed modes and eigenvectors revealing the detected modes' sample clusters.

In the proposed spectral framework, we attempt to compute the eigenspace of the kernel covariance matrices of the test and reference distributions. Assuming the Gaussian kernel with a properly chosen bandwidth, we prove that the eigenvalues and eigenvectors of the kernel covariance matrix will approximate the frequency and mean of the modes in a mixture distribution with well-separable components. Based on this result, we analyze the eigenspectrum of matrix $\Lambda_{\mathbf{X}|\eta\mathbf{Y}} = C_{\mathbf{X}} - \eta C_{\mathbf{Y}}$, i.e. the difference between the kernel covariance matrices of test $\mathbf{X}$ and reference $\mathbf{Y}$ data multiplied by coefficient $\eta \geq 1$. We demonstrate the application of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$'s eigendecomposition to identify the novel modes of test data $\mathbf{X}$ with an η-times higher frequency than in the reference data $\mathbf{Y}$. As a result, to quantify the mode-based novelty, we propose computing the entropy of the positive eigenvalues of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$, which we define to be the *Kernel-based Entropic Novelty (KEN)* score.

To compute the KEN score under high-dimensional kernel feature maps, e.g. the Gaussian kernel with an infinite-dimensional feature map, we develop a kernel-based method to compute the matrix's eigenvalues and eigenvectors. The proposed algorithm only requires the knowledge of pairwise kernel similarity scores between the observed $\mathbf{X}$ and $\mathbf{Y}$ samples and circumvents the computation challenges under high-dimensional kernel feature maps. Specifically, for a kernel function k, we show the η-differential kernel covariance matrix $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ shares the same eigenspectrum with the following matrix $K_{\mathbf{X}|\eta\mathbf{Y}}$ which we call the *η-differential kernel matrix*:

$$K_{\mathbf{X}|\eta\mathbf{Y}} = \begin{bmatrix} K_{\mathbf{XX}} & \sqrt{\eta}\, K_{\mathbf{XY}} \\ -\sqrt{\eta}\, K_{\mathbf{XY}}^{\top} & -\eta\, K_{\mathbf{YY}} \end{bmatrix},$$

where $K_{\mathbf{XX}}$, $K_{\mathbf{YY}}$, and $K_{\mathbf{XY}}$ denote the kernel matrices for $\mathbf{x}$ samples, $\mathbf{y}$ samples, and the cross kernel matrix between $\mathbf{x}$ and $\mathbf{y}$ samples, respectively. Also, while this matrix-based approach leads to the eigenvalue computation for the non-Hermitian matrix $K_{\mathbf{X}|\eta\mathbf{Y}}$, we show the application of Cholesky decomposition to reduce the problem to the eigendecomposition of a symmetric matrix, which can be handled more efficiently using standard linear algebra programming packages.

Finally, we present the numerical application of our proposed spectral method to several synthetic and real image datasets. For the synthetic experiments, we apply the novelty quantification and detection method to Gaussian mixture models and show the method can successfully count and identify the additional modes in the test distribution compared to the reference mixture model. In our experiments on real datasets, we apply the proposed method to identify the differently expressed sample clusters between standard image datasets. The numerical results suggest the methods' success in detecting the novel concepts present in the datasets. Furthermore, we apply the spectral method to detect the modes expressed with different frequencies by state-of-the-art generative modeling frameworks. The following is a summary of this work's main contributions:

- Proposing a kernel-based spectral method to analyze and quantify mode-based novelty across multi-modal distributions,
- Providing theoretical support for the novelty quantification method on mixture distributions with well-separable components,
- Developing a kernel method for computing the proxy mode centers and frequencies under high-dimensional kernel feature maps,
- Applying the spectral method to detect differently expressed modes between standard generative models.

2. Related Work

Fidelity and diversity evaluation of generative models. The evaluation of generative models has been studied in a large body of related works as surveyed in (Borji, 2022). The literature has proposed several metrics for the evaluation of the model's distance to the data distribution (Heusel et al., 2017; Bińkowski et al., 2018), quality, and diversity (Sajjadi et al., 2018; Kynkäänniemi et al., 2019; Naeem et al., 2020; Jalali et al., 2023; Dan Friedman & Dieng, 2023). Except the reference (Jalali et al., 2023), these works do not focus on mixture distributions. Also, our analysis concerns novelty evaluation between two distributions, different from the diversity assessment task addressed by Jalali et al. (2023).

Also, (Stein et al., 2023; Kynkäänniemi et al., 2023) have demonstrated that standard score-based evaluation methods may lead to a biased evaluation due to the choice of Inception-V3 embedding commonly used for image-based generative models. Stein et al. (2023) empirically show the less biased evaluation results using DINOv2 embedding. We note the similar importance of the selection of embedding in the results of the spectral KEN approach. We also highlight that the spectral method for evaluating KEN score results in an interpretable evaluation by identifying novel sample clusters between the test and reference distributions, whose relevance can be investigated by the evaluator.

Generalization evaluation of generative models. Several related works aim to measure the generalizability of generative models from training to test data. Alaa et al. (2022) use the percentage of authenticity to measure the likelihood of generated data copying the training data. Meehan et al. (2020) analyze training data-copying tendency by comparing the average distance to the closest training and test samples. Jiralerspong et al. (2023) examine overfitting by comparing the likelihoods based on training and test set. We note that the novelty evaluation task considered in our work is different from the generalizability assessment performed in these works, because our definition of mode-based novelty puts more emphasis on out-of-distribution modes not existing in the reference dataset. Also, we note that the reference dataset in out analysis may not be the training set of generative models, resulting in a different task from a training-to-test generalization evaluation.

Sample rarity and likelihood divergence. Han et al. (2023) empirically show that rare samples are far from the reference data in the feature space. They propose the rarity score as the nearest-neighbor distance to measure the uncommonness of image samples. Also, Jiralerspong et al. (2023) measure the difference in likelihood of generated distribution to the training and another reference dataset. They propose measuring the likelihood divergence and interpret novelty as low memorization of training samples. We note that both these evaluations lead to sample-based scores aiming to measure the uncommonness of a single data point. On the other hand, our proposed novelty evaluation is a distribution-based evaluation where we aim to measure the overall mode-based novelty of a model compared to a reference distribution.

3. Preliminaries

3.1. Novelty Evaluation of Generative Models

Consider a generator function $G : \mathbb{R}^r \to \mathbb{R}^d$ mapping an r-dimensional latent vector $\mathbf{Z}$ to $G(\mathbf{Z})$ which is aimed to be a real-like sample mimicking the data distribution P_{data}. Here $\mathbf{Z}$ is drawn according to a known distribution, e.g. an isotropic Gaussian $\mathcal{N}(\mathbf{0}, \sigma^2 I)$. However, the probability distribution of random vector $G(\mathbf{Z})$ could be challenging to compute for a neural network G. The goal in the evaluation of generative model $P_{G(\mathbf{Z})}$ is to quantify and estimate a desired property of its generated samples, e.g. quality and diversity, from n independently generated samples $G(\mathbf{z}_1), \dots, G(\mathbf{z}_n)$.

In this work, we focus on the evaluation of novelty in the generated data compared to a reference distribution Q for a random d-dimensional vector $\mathbf{Y}$. We assume we have access to m samples in $\{\mathbf{y}_1, \dots, \mathbf{y}_m\}$ drawn independently from Q. Also, for brevity, we denote the generative model G's generated data by $\mathbf{x}_i = G(\mathbf{z}_i)$ for every $1 \le i \le n$. Therefore, our aim is to quantify the novelty of generated dataset $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ compared to reference dataset $\{\mathbf{y}_1, \dots, \mathbf{y}_m\}$.

In our theoretical analysis, we assume the generated samples follow a multi-modal distribution. We use $P = \sum_{i=1}^k \omega_i P_i$ to represent a k-modal mixture distribution where every component P_i has frequency ω_i. Note that $[\omega_1, \dots, \omega_k]$ represent a probability model on the k modes in P satisfying $\omega_i \ge 0$ for every i and $\sum_{i=1}^k \omega_i = 1$. We assume each component P_i has σ_i^2-bounded total variance, defined as the trace of P_i's covariance matrix, meaning that given its mean vector $\boldsymbol{\mu}_i$, we have $\mathbb{E}_{\mathbf{X}\sim P_i}\left[\left\|\mathbf{X} - \boldsymbol{\mu}_i\right\|_2^2\right] \le \sigma_i^2$.

3.2. Kernel Function and Kernel Covariance Matrix

Consider a kernel function $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ mapping every two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ to a similarity score $k(\mathbf{x}, \mathbf{y})$ satisfying the positive semi-definite (PSD) property, i.e. the kernel matrix $K = \left[k(\mathbf{x}_i, \mathbf{x}_j)\right]_{n\times n}$ is a symmteric PSD matrix for every selection of input vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$. In this paper, we commonly suppose a normalized Gaussian kernel $k_{G(\sigma)}$ with bandwidth parameter σ defined as

$$k_{G(\sigma)}(\mathbf{x}, \mathbf{y}) := \exp\Big(\frac{-\|\mathbf{x} - \mathbf{y}\|_2^2}{2\sigma^2}\Big).$$

We remark that the PSD property of a kernel function k is equivalent to the existence of a feature map $\phi : \mathbb{R}^d \to \mathbb{R}^s$ such that for every input vectors $\mathbf{x}, \mathbf{y}$ we have $k\big(\mathbf{x}, \mathbf{y}\big) = \big\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \big\rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^s$. Also, we call a kernel function k normalized if for every $\mathbf{x} \in \mathbb{R}^d$ $k(\mathbf{x}, \mathbf{x}) = 1$, e.g. in the defined Gaussian kernel.

Given a distribution P with probability density function $p(\mathbf{x})$ on $\mathbf{X} \in \mathbb{R}^d$, we define the kernel covariance matrix according to kernel k with feature map ϕ as

$$C_{\mathbf{X}} := \mathbb{E}_{\mathbf{X}\sim P}\left[\phi(\mathbf{X})\phi(\mathbf{X})^\top\right] = \int p(\mathbf{x})\phi(\mathbf{x})\phi(\mathbf{x})^\top \mathrm{d}\mathbf{x}.$$

Using the empirical distribution $\widehat{P}_n$ of n samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ the kernel covariance matrix will be

$$\widehat{C}_{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n \phi(\mathbf{x}_i)\phi(\mathbf{x}_i)^\top,$$

which can be written as $\widehat{C}_{\mathbf{X}} = \frac{1}{n}\Phi_{\mathbf{X}}\Phi_{\mathbf{X}}^\top$ that $\Phi_{\mathbf{X}}$ is an $n \times s$ matrix with every i-th row being $\phi(\mathbf{x}_i)$.

Proposition 1. *Using the above definitions, $\widehat{C}_{\mathbf{X}}$ shares the same eigenvalues with the $n \times n$ normalized kernel matrix $\frac{1}{n}\left[k(\mathbf{x}_i, \mathbf{x}_j)\right]_{n\times n}$ with every (i, j)th entry being $\frac{1}{n}k(\mathbf{x}_i, \mathbf{x}_j)$. Therefore, assuming a normalized kernel function, the eigenvalues of $\widehat{C}_{\mathbf{X}}$ are non-negative and sum up to 1.*

4. A Spectral Approach to Novelty Evaluation for Mixture Models

In this section, we propose a spectral approach to the novelty evaluation of a generated $\mathbf{X}$ with mixture distribution

$P = \sum_{i=1}^{k} \omega_i P_i$ in comparison to a reference $\mathbf{Y}$ distributed according to mixture model $Q = \sum_{i=1}^{t} \gamma_i Q_i$. In what follows, we first define and intuitively explain the proposed novelty evaluation score, and later in this section, we will provide a theoretical analysis of the framework in a setting where the mixture models consist of well-separable components.

To define the proposed novelty score, we focus on the kernel covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, denoted by $C_{\mathbf{X}}$ and $C_{\mathbf{Y}}$, respectively. We will show in this section that under a Gaussian kernel with proper bandwidth, $C_{\mathbf{X}}$ and $C_{\mathbf{Y}}$ will contain the information of the modes in their eigendecomposition where the eigenvalues can be interpreted as the mode frequencies. Here, given a parameter $\eta \geq 1$, we define the *η-differential kernel covariance matrix* $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ as follows:

$$\Lambda_{\mathbf{X}|\eta\mathbf{Y}} \;:=\; C_{\mathbf{X}} - \eta C_{\mathbf{Y}}. \tag{1}$$

Note that if the components $P_1, \ldots, P_r$ are shared between P and Q, they will get canceled in the calculation of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ and thus will not result in a positive eigenvalue in the eigendecomposition of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ unless their frequency in P is greater than η-times their frequency in Q. This shows how we can, loosely speaking, "subtract Q from P" by subtracting their kernel covariance matrices in the above definition. Here, the hyperparameter $\eta \geq 1$ controls how much more frequent a mode in P must be compared to the corresponding mode in Q in order to be taken into account.

Therefore, we use the positive eigenvalues of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ to approximate the relative frequencies of the modes of P that are expressed at least η-times more frequently than in Q. This allows us to define the following entropic score to quantify the novelty of $\mathbf{X}$ with respect to $\mathbf{Y}$.

Definition 1. *Consider the positive eigenvalues $\lambda_1, \ldots, \lambda_{k'} > 0$ of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}} = C_{\mathbf{X}} - \eta C_{\mathbf{Y}}$ and let $S = \sum_{i=1}^{k'} \lambda_i$. The Kernel Entropic Novelty (KEN) score is*

$$\mathrm{KEN}_{\eta}(\mathbf{X}|\mathbf{Y}) \;:=\; \sum_{i=1}^{k'} \lambda_i \log \frac{S}{\lambda_i}. \tag{2}$$

Intuitively, λ_i/S is the relative frequency of the i-th novel mode in P, and the entropy of the novel mode distribution is $\sum_i (\lambda_i/S)\log(S/\lambda_i)$. The entropy is multiplied by S in the definition of $\mathrm{KEN}_{\eta}(\mathbf{X}|\mathbf{Y})$ for two reasons. First, the amount of novelty should not only increase with the entropy (or diversity) of the novel modes, but also increase with the total frequency S of those modes. Second, this allows us to interpret $\mathrm{KEN}_{\eta}(\mathbf{X}|\mathbf{Y})$ as the conditional entropy of the information of the mode of $\mathbf{X}$ given the dataset of $\mathbf{Y}$. Later in this section, we will present theoretical results that justify the above informal discussions, and the interpretation of $\mathrm{KEN}_{\eta}(\mathbf{X}|\mathbf{Y})$ as a conditional entropy.

4.1. Theoretical Analysis of the Proposed Spectral Novelty Evaluation

To theoretically analyze the proposed spectral method, we first focus on a single distribution P with well-separable modes and show a relationship between the modes of P and the eigendecomposition of its kernel covariance matrix. We defer the proof of the theoretical results to the Appendix.

Theorem 1. *Suppose that every component P_i of a mixture distribution $P = \sum_{i=1}^{k} \omega_i P_i$ has mean vector $\boldsymbol{\mu}_i$ and bounded total variance $\mathbb{E}_{\mathbf{X}\sim P_i}\big[\|\mathbf{X} - \boldsymbol{\mu}_i\|_2^2\big] \leq \sigma_i^2$. Assume that $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_k$ are sorted in a descending order. Then, the top k eigenvalues $\lambda_1 \geq \cdots \geq \lambda_k$ of the kernel covariance matrix $C_{\mathbf{X}}$ according to a Gaussian kernel with bandwidth σ will satisfy:*

$$\sum_{i=1}^{k}\big(\lambda_i - \omega_i\big)^2 \;\leq 4\sum_{i=1}^{k}\omega_i\frac{\sigma_i^2}{\sigma^2} + 16\sum_{i=2}^{k}\sum_{j=1}^{i-1}\omega_i \exp\Big(\frac{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big).$$

The above theorem shows that if the modes of the mixture distribution are well-separable, meaning that $\min_{1\leq i\neq j\leq k}\frac{\|\boldsymbol{\mu_i}-\boldsymbol{\mu_j}\|_2}{\sigma} \gg 1$ while the total variance of the components satisfies $\frac{\sigma_i}{\sigma} \ll 1$, then the eigendecomposition of the Gaussian kernel covariance matrix can reveal the mode frequencies via the principal eigenvalues.

Given the interpretation provided by Theorem 1, the positive eigenvalues of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}} = C_{\mathbf{X}} - \eta C_{\mathbf{Y}}$ will correspond to the modes of $\mathbf{X}$ that have a frequency at least η times higher than the frequency of that mode in $\mathbf{Y}$. Therefore, the positive eigenvalues of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ can be used to quantify the novelty of $\mathbf{X}$ compared to $\mathbf{Y}$. The following theorem formalizes this intuition and shows how $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$'s positive eigenvalues explain the novelty of $\mathbf{X}$'s modes.

Theorem 2. *Consider multi-modal random vectors $\mathbf{X} \sim \sum_{i=1}^{k}\omega_i P_i$ and $\mathbf{Y} \sim \sum_{i=1}^{k}\gamma_i Q_i$, where $\omega_1 - \eta\gamma_1 \geq \cdots \geq \omega_k - \eta\gamma_k$. Suppose the corresponding mode to every P_i with mean $\boldsymbol{\mu}_i$ is Q_i with mean $\boldsymbol{\mu}'_i = \boldsymbol{\mu}_i + \boldsymbol{\delta}_i$. Then, assuming that for every i, both Q_i and P_i have total variance bounded by σ_i^2, the positive eigenvalues $\lambda_1 \geq \ldots \geq \lambda_{k'} > 0$ of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ satisfy (letting $\lambda_i = 0$ if $i > k'$)*

$$\begin{aligned}&\sum_{i=1}^{k}\Big(\lambda_i - \max\big\{\omega_i - \eta\gamma_i, 0\big\}\Big)^2 \\ \leq\; & 8\sum_{i=1}^{k}\Big[\omega_i\frac{\|\boldsymbol{\delta}_i\|_2^2}{\sigma^2} + \big(\omega_i + \eta^2\gamma_i\big)\frac{\sigma_i^2}{\sigma^2}\Big] \\ & + 16(1+\eta)\sum_{i=2}^{k}\sum_{j=1}^{i-1}(\omega_i + \eta\gamma_i)\exp\Big(\frac{-\|\boldsymbol{\mu}'_i - \boldsymbol{\mu}'_j\|_2^2}{\sigma^2}\Big).\end{aligned}$$

Based on the above theorem, the principal positive eigenvalues of the defined matrix $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$ show the extra frequencies of the modes with a more dominant presence in $\mathbf{X}$. As we increase the value of η, we require a stronger presence of $\mathbf{X}$'s modes to count them as a novel mode. In the limit case where $\eta \to +\infty$, we require a complete absence of an $\mathbf{X}$'s mode in $\mathbf{Y}$ to call it novel.

Finally, note that $\mathrm{KEN}_\eta(\mathbf{X}|\mathbf{Y})$ can be interpreted as the conditional entropy of the information of the mode of $\mathbf{X}$ given the dataset of $\mathbf{Y}$. More precisely, if $\eta = 1$ this score is the conditional entropy $H(X_{\mathrm{mode}}|Y_{\mathrm{adv}})$, where $X_{\mathrm{mode}} \in \{1, \ldots, k\}$ is the mode cluster variable of $\mathbf{X}$ (the index of the mode $\mathbf{X}$ belongs to), and Y_{adv} represents the knowledge of an adversary who knows the dataset of $\mathbf{Y}$ and wants to predict X_{mode}. If the random sample $\mathbf{X}$ is also found in the dataset of $\mathbf{Y}$, then the adversary will know the mode of $\mathbf{X}$ and predicts $Y_{\mathrm{adv}} = X_{\mathrm{mode}}$ accurately; otherwise the adversary knows nothing about X_{mode}, and outputs $Y_{\mathrm{adv}} = e$ as an erasure symbol denoting the lack of information.

Under the setting in Theorem 2 for $\eta = 1$, $\lambda_i = \max\{\omega_i - \gamma_i, 0\}$, take $\mathbb{P}(Y_{\mathrm{adv}} = i|X_{\mathrm{mode}} = i) = \min\{\gamma_i/\omega_i, 1\}$ (otherwise $Y_{\mathrm{adv}} = e$) since among the samples of $\mathbf{X}$ with mode i, at most a portion γ_i/ω_i are also samples of $\mathbf{Y}$ (if sizes of the two datasets are equal). We have $\mathbb{P}(Y_{\mathrm{adv}} = e) = \sum_{i=1}^{t} \omega_i \max\{1 - \gamma_i/\omega_i, 0\} = S$. Since $H(X_{\mathrm{mode}}|Y_{\mathrm{adv}} = i) = 0$, we have

$$\begin{aligned} H(X_{\mathrm{mode}}|Y_{\mathrm{adv}}) &= \mathbb{P}(Y_{\mathrm{adv}} = e) H(X_{\mathrm{mode}}|Y_{\mathrm{adv}} = e) \\ &= S \sum_{i=1}^{t} \frac{\lambda_i}{S} \log \frac{S}{\lambda_i} \;=\; \mathrm{KEN}_1(\mathbf{X}|\mathbf{Y}). \end{aligned}$$

For a general η, $\mathrm{KEN}_\eta(\mathbf{X}|\mathbf{Y})$ can be interpreted as $H(X_{\mathrm{mode}}|Y_{\mathrm{adv}})$ if each sample of $\mathbf{Y}$ allows the adversary to learn η samples of $\mathbf{X}$ belonging to the same mode as $\mathbf{Y}$, resulting in $\mathbb{P}(Y_{\mathrm{adv}} = i \,|\, X_{\mathrm{mode}} = i) = \min\{\eta\gamma_i/\omega_i, 1\}$. The next section shows how we can use the kernel trick to compute the KEN score from the pairwise similarity scores between the observed test and reference samples.

5. Computation of the KEN Novelty Score

Since the KEN score is characterized using the difference $C_{\mathbf{X}} - \eta C_{\mathbf{Y}}$ of kernel covariance matrices, the computation of this score will be challenging in the kernel feature space under a high-dimensional kernel feature map. Specifically, the kernel feature map for a Gaussian kernel is infinitely high-dimensional. Our next theorem reduces the eigendecomposition task to the differential kernel matrix based on only the kernel similarity scores of empirical samples.

Theorem 3. *Suppose we observed empirical samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from test distribution P and $\mathbf{y}_1, \ldots, \mathbf{y}_m$ from reference distribution Q. Then, the difference of empirical kernel covariance matrices $\widehat{\Lambda}_{\mathbf{X}|\eta\mathbf{Y}} = \widehat{C}_{\mathbf{X}} - \eta\widehat{C}_{\mathbf{Y}}$ shares the same positive eigenvalues with the following matrix, which we call the η-*differential kernel matrix*:*

$$K_{\mathbf{X}|\eta\mathbf{Y}} := \begin{bmatrix} K_{\mathbf{XX}} & \sqrt{\eta}\, K_{\mathbf{XY}} \\ -\sqrt{\eta}\, K_{\mathbf{XY}}^\top & -\eta\, K_{\mathbf{YY}} \end{bmatrix} \tag{3}$$

In the above, $K_{\mathbf{XX}} = \frac{1}{n}\big[k(\mathbf{x}_i, \mathbf{x}_j)\big]_{n\times n}$ and $K_{\mathbf{YY}} = \frac{1}{m}\big[k(\mathbf{y}_i, \mathbf{y}_j)\big]_{m\times m}$ are the kernel similarity matrices for observed $\mathbf{X}$ and $\mathbf{Y}$, samples respectively, and $K_{\mathbf{XY}} = \frac{1}{\sqrt{nm}}\big[k(\mathbf{x}_i, \mathbf{y}_j)\big]_{n\times m}$ is the $n \times m$ cross-kernel matrix between observed $\mathbf{X}, \mathbf{Y}$ samples.

Theorem 3 simplifies the eigendecomposition of the matrix $\widehat{\Lambda}_{\mathbf{X}|\eta\mathbf{Y}}$ to the $(n+m)\times(n+m)$ kernel-based matrix $K_{\mathbf{X}|\eta\mathbf{Y}}$. We remark that each eigenvector $\mathbf{v}_i \in \mathbb{R}^{m+n}$ of this matrix contains the expected inner-product of empirical $\mathbf{X}$ and $\mathbf{Y}$ data with the ith detected mode, which can be utilized to rank the samples based on their likelihood of belonging to the ith identified novel cluster. Due to the sign ambiguity for each eigendirection $\mathbf{v}_i$, we multiply computed eigenvector $\mathbf{v}_i$ by the sign of sum of its first n entries, $\mathrm{sgn}\big(\sum_{j=1}^{n} v_{i,j}\big)$, to prefer a positive score for test $\mathbf{x}_1, \ldots, \mathbf{x}_n$ samples.

While the discussed eigendecomposition can be addressed via $O\big((n+m)^3\big)$ computations, $K_{\mathbf{X}|\eta\mathbf{Y}}$ is a non-Hermitian matrix, for which standard Hermitian matrix-based algorithms do not apply. In the following theorem, we apply Cholesky decomposition to reduce the task to an eigenvalue computation for a symmetric matrix.

Theorem 4. *In the setting of Theorem 3, define the following joint kernel matrix:*

Algorithm 1 Computation of KEN & novel mode centers

Input: Sample sets $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$, parameter $\eta > 0$, Gaussian kernel bandwidth σ

Compute matrices $K_{\mathbf{XX}} = \frac{1}{n}\big[k(\mathbf{x}_i, \mathbf{x}_j)\big]_{n\times n}$, $K_{\mathbf{YY}} = \frac{1}{m}\big[k(\mathbf{y}_i, \mathbf{y}_j)\big]_{m\times m}$, $K_{\mathbf{XY}} = \frac{1}{\sqrt{mn}}\big[k(\mathbf{x}_i, \mathbf{y}_j)\big]_{n\times m}$

Apply Cholesky decomposition to compute upper-triangular $V \in \mathbb{R}^{(n+m)\times(n+m)}$ such that

$$V^\top V \;=\; \begin{bmatrix} K_{\mathbf{XX}} & \sqrt{\eta}\, K_{\mathbf{XY}} \\ \sqrt{\eta}\, K_{\mathbf{XY}}^\top & \eta\, K_{\mathbf{YY}} \end{bmatrix}$$

Compute $\Gamma = V\,\mathrm{diag}([\underbrace{+1, \ldots, +1}_{n \text{ times}}, \underbrace{-1, \ldots, -1}_{m \text{ times}}])V^\top$

Perform eigendecomposition to get $\Gamma = U^\top \mathrm{diag}(\boldsymbol{\lambda})U$

Find the positive eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{k'} > 0$ and the corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k'}$

Set $\mathbf{u}_i \leftarrow \mathrm{sgn}(\sum_{j=1}^{n} u_{i,j}) \cdot \mathbf{u}_i$ for $i = 1, \ldots, k'$

Output: KEN-score $= \sum_{i=1}^{k'} \lambda_i \log \frac{\sum_{j=1}^{k'} \lambda_j}{\lambda_i}$ and eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k'}$.

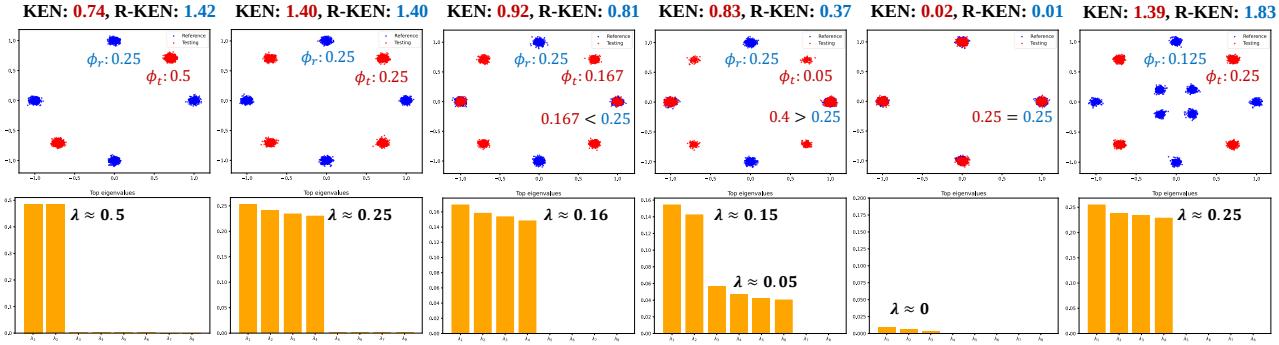

Figure 1. Experimental results on synthetic Gaussian mixture distributions including KEN and R-KEN (Reversed-KEN) scores, and principal eigenvalues of the differential kernel covariance matrix $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$. **Top row:** Reference (in blue) and test (in red) samples with ϕ_t, ϕ_r denoting the test and reference modes' frequency. **Bottom row:** Positive eigenvalues of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}}$.

$$K_{\mathbf{X},\eta\mathbf{Y}} := \begin{bmatrix} K_{\mathbf{XX}} & \sqrt{\eta}\, K_{\mathbf{XY}} \\ \sqrt{\eta}\, K_{\mathbf{XY}}^{\top} & \eta\, K_{\mathbf{YY}} \end{bmatrix}.$$

Consider the Cholesky decomposition of the above PSD matrix satisfying $K_{\mathbf{X},\eta\mathbf{Y}} = V^{\top}V$ for upper-triangular matrix $V \in \mathbb{R}^{(n+m)\times(n+m)}$. Then, $\widehat{\Lambda}_{\mathbf{X}|\eta\mathbf{Y}}$ shares the same non-zero eigenvalues with the symmetric matrix $VDV^{\top}$ where D is a $(n+m)\times(n+m)$ diagonal matrix with diagonal entries in $[\underbrace{+1,\ldots,+1}_{n \text{ times}},\underbrace{-1,\ldots,-1}_{m \text{ times}}]$.

Based on the above theorem, we propose Algorithm 1 to compute the KEN score and find the eigendirections corresponding to the detected novel modes. We note that the eigendecomposition task in the algorithm reduces to the spectral decomposition of a symmetric matrix that can be handled more efficiently than the eigenvalue computation for a general non-symmetric matrix.

Finally, note that each computed eigenvector $\mathbf{u}_i \in \mathbb{R}^{n+m}$ in Algorithm 1 corresponds to the function $\widetilde{u}_i : \mathbb{R}^d \to \mathbb{R}$,

$$\widetilde{u}_i(\mathbf{x}) = \sum_{j=1}^{n} u_{i,j} k(\mathbf{x}_j, \mathbf{x}) + \sum_{s=1}^{m} u_{i,s+n} k(\mathbf{y}_s, \mathbf{x}),$$

where $u_{i,j}$ stands for the j-th entry of $\mathbf{u}_i$. The above function's output $\widetilde{u}_i(\mathbf{x})$ can be viewed as the data point $\mathbf{x}$'s score of belonging to the identified i-th cluster. Therefore, we include Step 7 in Algorithm 1, which multiplies the computed eigenvector $\mathbf{u}_i$ with $+1$ or -1, to prefer a non-negative score for test data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in the novelty evaluation task.

6. Numerical Results

6.1. Experimental Setup

Datasets. We performed experiments on the following image datasets: 1) CIFAR-10 (Krizhevsky et al., 2009) with 60k images of 10 classes, 2) ImageNet-1K (Deng et al., 2009) with 1.4 million images of 1000 classes, containing 20k dog images from 120 different dog breeds, 3) CelebA (Liu et al., 2015) with 200k face images of celebrities, 4) FFHQ (Karras et al., 2019) with 70k human-face images, 5) AFHQ (Choi et al., 2020) with 15k animal-face images of dogs, cats, and wildlife. The AFHQ-dog subset has 5k images from 8 dog breeds. 6) Wildlife dataset (Mehta, 2023) with 2k wild animal images.

Pre-trained generative models and neural nets for feature extraction: We used the following embeddings in our experiments: 1) pre-trained Inception-V3 (Szegedy et al., 2016) which is the standard in FID and IS scores. 2) DINOv2 (Oquab et al., 2023) suggested by Stein et al. (2023) to reduce the biases in ImageNet-based Inception-V3 embedding, 3) CLIP (Radford et al., 2021) suggested by Kynkäänniemi et al. (2023) to lessen the inductive biases of Inception-V3 embedding. For a fair comparison between the tested image-based generative models, we downloaded the pre-trained models from the StudioGAN (Kang et al., 2023) and (Stein et al., 2023)'s GitHub repositories.

Bandwidth parameter σ and sample size: Similar to (Jalali et al., 2023), we chose the kernel bandwidth to be the smallest σ satisfying variance < 0.01. In our experiments, we observed $\sigma \in [10, 15]$ could satisfy this requirement for all the tested image data with the Inception-V3 embedding. In the case of synthetic Gaussian mixtures, we used $\sigma = 0.5$. In our experiments, we used $m, n = 5000$ sample size for the test and reference data.

6.2. Numerical Results on Synthetic Gaussian Mixtures

First, we tested the proposed methodology on Gaussian mixture models (GMMs) as shown in Figure 1. The experiments use the standard setting of 2-dimensional Gaussian mixtures in (Gulrajani et al., 2017). We show the samples from the reference distribution (in blue) with a 4-component GMM where the components are centered at [0, 1], [1, 0], [0, -1], [-1, 0]. The generated data in the test distribution

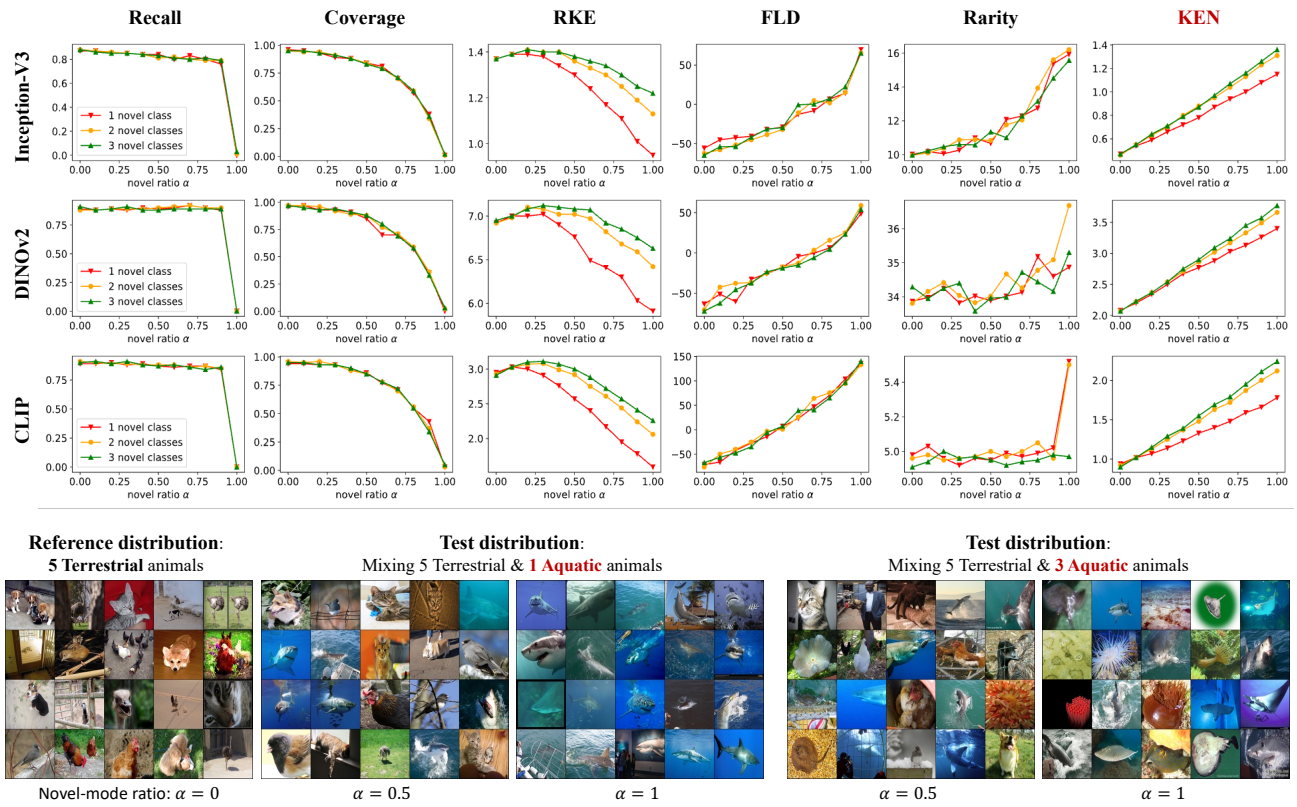

Figure 2. **Top 3 rows:** Trends of baseline and KEN scores in evaluating novel yet less-diverse distributions with Inception-V3, DINOv2 and CLIP embeddings. **Bottom:** ImageNet-1K Samples from reference and test distributions. Reference modes: 5 terrestrial animals. Novel modes: 1-3 aquatic lives. α is the ratio of novel modes in testing distribution. $\alpha = 0, 1$ represents pure reference and novel distributions, respectively.

(in red) follow a Gaussian mixture in all the experiments, where we center the novel modes (unexpressed in the reference) at $[\pm 0.7, \pm 0.7]$ and center the shared modes at the same component-means of the reference distribution. In the experiments, we chose parameter $\eta = 1$ for the KEN evaluation.

Based on the KEN scores and eigenspectrum of the η-differential kernel matrix reported in Figure 1, our proposed spectral method successfully identifies the novel modes, and the KEN scores correlate with the novel modes' number and frequencies. We highlight the following trends in the evaluated KEN scores:

1. More novel modes result in a greater KEN score. The first two columns of Figure 1 illustrate that adding two novel modes to the test distribution increases the KEN score from 0.74 to 1.40. The bar plots of the differential kernel matrix's eigenvalues also show two extra principal eigenvalues approximating the frequencies of novel modes.

2. Transferring weight from novel to common modes decreases KEN score. Columns 3-5 in Figure 1 show the effects of overlapping modes on KEN score. In Column 3, the test distribution has six components with uniform frequencies of $1/6$, of which two modes are centered at the same points as the reference modes with frequency $1/4$. The KEN score decreased from 1.40 to 0.92, and we observed only four principal eigenvalues in the differential kernel matrix. Also, when we increased the frequencies of the common modes from 0.25 to 0.4, as shown in Column 4, we could observe 6 outstanding eigenvalues, from which two of them approximate the difference of common mode frequencies. Moreover, under two identical distributions, the KEN score was nearly 0.

3. KEN does not behave symmetrically between the reference and test distributions. In our experiments, we also measured the KEN score of the reference distribution with respect to the test distribution, which we call *Reverse KEN (R-KEN)*. We observed that the KEN and R-KEN scores could behave differently, and the roles of test and reference distributions were different. Regarding KEN and R-KEN's mismatch, when we included four extra *reference* modes in the last column, the KEN score did not considerably change

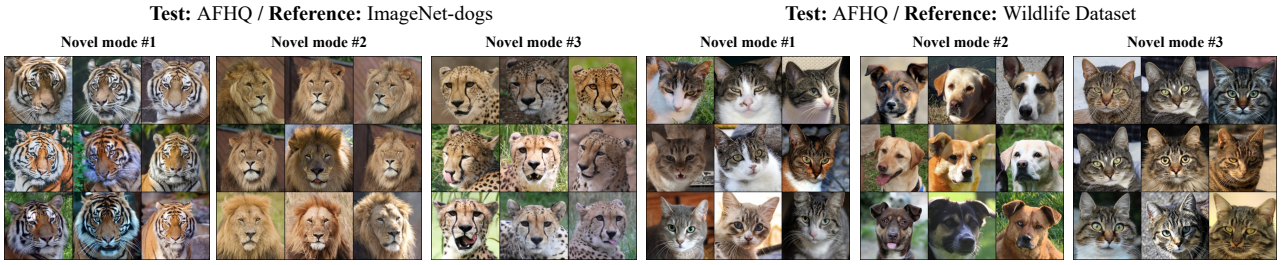

Figure 3. Identified top-3 novel modes between image datasets: (Left-half) AFHQ w.r.t. ImageNet-dogs, (Right-half) AFHQ w.r.t. Wildlife. Inception-V3 embedding is used. Shown samples are the test data with the maximum entry values on the top three principal eigenvectors of the differential kernel matrix $K_{\mathbf{X}|\eta\mathbf{Y}}$ defined in (3).

(1.40 vs. 1.39) while the R-KEN value jumped from 1.40 to 1.83.

6.3. Novelty vs. Diversity Evaluation via KEN and Baseline Metrics

The novelty and diversity evaluation criteria may not align under certain conditions. To test our proposed method and the existing scores' capability to capture novelty under less diversity, we designed experimental settings where the test distribution possessed less diversity while containing novel modes compared to the reference distribution. The baseline diversity-based scores we attempted in the experiment were improved Recall (Kynkäänniemi et al., 2019), Coverage (Naeem et al., 2020), and RKE (Jalali et al., 2023). We also evaluated the sample divergence: FLD (Jiralerspong et al., 2023) and sample Rarity (Han et al., 2023), which are proposed to assess sample-based novelty. We extract images from 5 terrestrial animal classes in ImageNet-1K to form the reference distribution P_r and images from 1-3 aquatic life classes to form the novel distribution P_n. To simulate different novelty ratios, we mixed the two distributions with a ratio parameter $0 \leq \alpha \leq 1$ to form the test distribution as $P_t = \alpha P_n + (1-\alpha)P_r$.

As shown in Figure 2, all the diversity-based baseline metrics decreased under a larger α, i.e., as the test distribution P_t becomes closer to the novel distribution P_n. The observation can be interpreted as the diversity and novelty levels change in opposite directions in this experiment. The proposed KEN score and baseline FLD and Rarity scores could capture the higher novelty and increased with α. On the other hand, when we increased the number of novel modes from 1 to 3 aquatic animals, the sample-based FLD and Rarity scores did not change significantly, while our proposed KEN score could capture the extra novel modes and grew with the number of novel modes. This experiment shows the *distribution-based* novelty evaluation by the KEN score vs. the *sample-based* novelty evaluation by FLD and Rarity.

6.4. Numerical Results on Real/Generated Image Data

We evaluated the KEN score and visualized identified novel modes for the real image dataset and sample sets generated by widely-used generative models. For the identification of samples belonging to the detected novel modes, we followed Algorithm 1 to obtain the eigenvectors corresponding to the top eigenvalues. Every eigenvector $\mathbf{u}_i$ is an $(m+n)$-dimensional vector where the entries (sample indices) with significant values greater than a threshold ρ are clustered as mode i. In our visualization, we show top-r images with the maximum entry value on the shown top eigenvectors as the top novel modes.

Novel modes between real datasets. Based on the proposed method, we visualized the novel modes samples' across content-similar datasets (AFHQ, ImageNet-dogs) and (AFHQ, Wildlife). Figure 3 visualizes the identified samples from the top three novel modes (principal eigenvectors). We observed that the detected modes exhibit semantically meaningful picture types of novel wildlife types missing in the ImageNet-dogs samples and novel docile "cat" and "dog" types compared to the Wildlife dataset. We did not find such samples when searching for these image types in the reference datasets. We postpone the presentation of similar results on other dataset pairs and CLIP and DINOv2 embeddings to the Appendix.

Novel modes between standard generative models. We analyzed FFHQ-trained models: GAN-based InsGen, StyleGAN2, StyleGAN-XL, diffusion-based LDM, and VAE-based VDVAE. Figure 4 illustrates samples from the detected top novel mode between different pairs of generative models. For example, we observed that InsGen has more novelty in "people wearing sunglasses" than LDM, and StyleGAN2 has more novelty in "kids" than VDVAE. We present and discuss more visualizations for other pairs of generative models in Appendix A.2. According to Table 1, considering the averaged KEN scores over all $\binom{5}{2}$ pairs, InsGen obtained the maximum averaged-KEN among the tested generative models in the FFHQ case.

Detection of missing sample types of generative models.

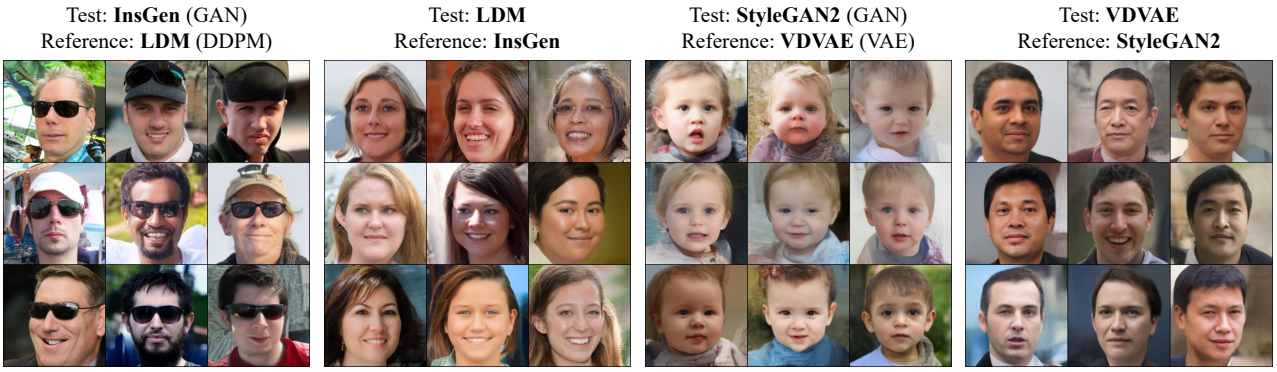

Figure 4. Identified top novel modes between FFHQ-trained generative models. Inception-V3 embedding is used.

Table 1. FFHQ-trained generative models' pairwise KEN score. Inception-V3 embedding is used.

Generative Models (Test Models)	Reference Models					Avg. KEN
	InsGen	LDM	StyleGAN2	VDVAE	StyleGAN-XL	
InsGen (Yang et al., 2021)	-	1.26	1.18	1.87	1.17	1.37
LDM (Rombach et al., 2021)	1.09	-	1.14	1.59	1.08	1.23
StyleGAN2 (Karras et al., 2019)	1.12	1.26	-	1.76	1.18	1.33
VDVAE (Child, 2020)	0.96	0.91	0.94	-	0.95	0.94
StyleGAN-XL (Sauer et al., 2022)	1.16	1.24	1.19	1.83	-	1.36

To detect missing modes of the generators, we used the larger value $\eta = 10$ for parameter η in the computation of $K_{\mathbf{X}|\eta\mathbf{Y}}$. For example, our experimental results suggest the sample types "Microphone", "Round hat", and "Black uniform hat" to be not well-expressed in LDM. The visualization of our numerical results is postponed to the Appendix. In the Appendix, we will also present the applications of our spectral method for conditionally generating novel-mode samples and benchmarking model fitness.

7. Conclusion

In this paper, we proposed a spectral method for the evaluation of the novel modes in a mixture distribution P which are expressed more frequently than in a reference distribution Q. We defined the KEN score to measure the entropy of the novel modes and tested the evaluation method on benchmark synthetic and image datasets. We note that our numerical evaluation focused on computer vision settings, and its extension to language models will be an interesting future direction. Also, characterizing tight statistical and computational complexity bounds for the novelty evaluation method will be a related topic for future exploration.

8. Limitations

Similar to other evaluation methods for image-based generative models, the results of KEN novelty evaluation are influenced by the choice of embedding, which may lead to biased results under ImageNet pre-trained models, such as the standard Inception-V3. Our work mainly focused on introducing and developing the kernel method for KEN novelty evaluation, and we leave a detailed analysis of the role of embedding in the novelty assessment, similar to the analysis in (Kynkäänniemi et al., 2023; Stein et al., 2023) for quality and diversity metrics, for future studies. Furthermore, the spectral algorithm for KEN evaluation requires eigendecomposition of an $(n+m) \times (n+m)$ kernel matrix for n, m test and reference samples, whose computational complexity $O\big((n+m)^3\big)$ will remain a barrier towards applying the framework to large sample sizes needed for large-scale datasets e.g. ImageNet. Exploring scalable extensions of the KEN framework is an interesting future direction.

Acknowledgements

The work of Farzan Farnia is partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China, Project 14209920, and is partially supported by a CUHK Direct Research Grant with CUHK Project No. 4055164. The work of Cheuk Ting Li is partially supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No.s: CUHK 24205621 (ECS), CUHK 14209823 (GRF)]. Also, the authors would like to thank the anonymous reviewers for their constructive feedback and useful suggestions.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Alaa, A., Van Breugel, B., Saveliev, E. S., and van der Schaar, M. How faithful is your synthetic data? sample-level metrics for evaluating and auditing generative models. In *International Conference on Machine Learning*, pp. 290–306. PMLR, 2022.

Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein gan, 2017.

Bińkowski, M., Sutherland, D. J., Arbel, M., and Gretton, A. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018.

Borji, A. Pros and cons of gan evaluation measures: New developments. *Computer Vision and Image Understanding*, 215:103329, 2022. ISSN 1077-3142.

Brock, A., Donahue, J., and Simonyan, K. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

Child, R. Very deep vaes generalize autoregressive models and can outperform them on images. *arXiv preprint arXiv:2011.10650*, 2020.

Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8188–8197, 2020.

Dan Friedman, D. and Dieng, A. B. The vendi score: A diversity evaluation metric for machine learning. *Transactions on machine learning research*, 2023.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

Han, J., Choi, H., Choi, Y., Kim, J., Ha, J.-W., and Choi, J. Rarity score : A new metric to evaluate the uncommonness of synthesized images. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=JTGimap_-F.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Hoffman, A. J. and Wielandt, H. W. The variation of the spectrum of a normal matrix. In *Selected Papers Of Alan J Hoffman: With Commentary*, pp. 118–120. World Scientific, 2003.

Jalali, M., Li, C. T., and Farnia, F. An information-theoretic evaluation of generative models in learning multi-modal distributions. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Jiralerspong, M., Bose, J., Gemp, I., Qin, C., Bachrach, Y., and Gidel, G. Feature likelihood score: Evaluating the generalization of generative models using samples. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=l2VKZkolT7.

Kang, M. and Park, J. Contragan: Contrastive learning for conditional image generation, 2021.

Kang, M., Shim, W., Cho, M., and Park, J. Rebooting acgan: Auxiliary classifier gans with stable training, 2021.

Kang, M., Shin, J., and Park, J. StudioGAN: A Taxonomy and Benchmark of GANs for Image Synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., and Aila, T. Training generative adversarial networks with limited data, 2020.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., and Aila, T. Alias-free generative adversarial networks. In *Proc. NeurIPS*, 2021.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., and Aila, T. *Improved Precision and Recall Metric for Assessing Generative Models*. Curran Associates Inc., Red Hook, NY, USA, 2019.

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., and Aila, T. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems*, 32, 2019.

Kynkäänniemi, T., Karras, T., Aittala, M., Aila, T., and Lehtinen, J. The role of imagenet classes in fréchet inception distance. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=4oXTQ6m_ws8.

Lim, J. H. and Ye, J. C. Geometric gan. *arXiv preprint arXiv:1705.02894*, 2017.

Liu, Z., Luo, P., Wang, X., and Tang, X. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pp. 3730–3738, 2015.

Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., and Paul Smolley, S. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pp. 2794–2802, 2017.

Marchesi, M. Megapixel size image creation using generative adversarial networks. *arXiv preprint arXiv:1706.00082*, 2017.

Meehan, C., Chaudhuri, K., and Dasgupta, S. A non-parametric test to detect data-copying in generative models. In *International Conference on Artificial Intelligence and Statistics*, 2020.

Mehta, A. Wildlife animals images, version 1, 2023. URL https://www.kaggle.com/datasets/anshulmehtakaggl/wildlife-animals-images.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks, 2018.

Naeem, M. F., Oh, S. J., Uh, Y., Choi, Y., and Yoo, J. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning*, pp. 7176–7185. PMLR, 2020.

Odena, A., Olah, C., and Shlens, J. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pp. 2642–2651. PMLR, 2017.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H. V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Howes, R., Huang, P.-Y., Xu, H., Sharma, V., Li, S.-W., Galuba, W., Rabbat, M., Assran, M., Ballas, N., Synnaeve, G., Misra, I., Jegou, H., Mairal, J., Labatut, P., Joulin, A., and Bojanowski, P. Dinov2: Learning robust visual features without supervision, 2023.

Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. Learning transferable visual models from natural language supervision, 2021.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models, 2021.

Sajjadi, M. S., Bachem, O., Lucic, M., Bousquet, O., and Gelly, S. Assessing generative models via precision and recall. *Advances in neural information processing systems*, 31, 2018.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., and Chen, X. Improved techniques for training GANs. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Sauer, A., Schwarz, K., and Geiger, A. Stylegan-xl: Scaling stylegan to large diverse datasets. In *ACM SIGGRAPH 2022 conference proceedings*, pp. 1–10, 2022.

Stein, G., Cresswell, J. C., Hosseinzadeh, R., Sui, Y., Ross, B. L., Villecroze, V., Liu, Z., Caterini, A. L., Taylor, J. E. T., and Loaiza-Ganem, G. Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models, 2023.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Wu, Y., Donahue, J., Balduzzi, D., Simonyan, K., and Lillicrap, T. Logan: Latent optimisation for generative adversarial networks. *arXiv preprint arXiv:1912.00953*, 2019.

Yang, C., Shen, Y., Xu, Y., and Zhou, B. Data-efficient instance generation from instance discrimination. *arXiv preprint arXiv:2106.04566*, 2021.

Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. Self-attention generative adversarial networks. In *International conference on machine learning*, pp. 7354–7363. PMLR, 2019.

A. Appendix

A.1. Proofs

A.1.1. Proof of Theorem 1

To prove the theorem, note that every mode variable $\mathbf{X}_i \sim P_i$ can be written as $\mathbf{X}_i = \boldsymbol{\mu}_i + \mathbf{V}_i$ where $\mathbf{V}_i$ is a zero-mean random vector satisfying a bounded second-order moment $\mathbb{E}\big[\|\mathbf{V}_i\|_2^2\big] \leq \sigma_i^2$. Then, we can decompose the kernel covariance matrix $C_{\mathbf{X}}$ into the following two terms:

$$C_{\mathbf{X}} \;=\; \sum_{i=1}^{k}\Big[\omega_i \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\Big] + \sum_{i=1}^{k}\Big[\omega_i\Big(\mathbb{E}\big[\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top\big] - \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\Big)\Big].$$

Therefore, we can write

$$\begin{aligned}
\Big\|C_{\mathbf{X}} - \sum_{i=1}^{k}\omega_i\phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\Big\|_F^2 &= \Big\|\sum_{i=1}^{k}\Big[\omega_i\Big(\mathbb{E}\big[\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top\big] - \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\Big)\Big]\Big\|_F^2 \\
&\overset{(a)}{\leq} \sum_{i=1}^{k}\Big[\omega_i\Big\|\mathbb{E}\big[\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top\big] - \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\Big\|_F^2\Big] \\
&= \sum_{i=1}^{k}\Big[\omega_i\Big\|\mathbb{E}\big[\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top - \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\big]\Big\|_F^2\Big] \\
&\overset{(b)}{\leq} \sum_{i=1}^{k}\omega_i\mathbb{E}\Big[\Big\|\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top - \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\Big\|_F^2\Big] \\
&\overset{(c)}{=} \sum_{i=1}^{k}\omega_i\mathbb{E}\Big[2 - 2\big(\phi(\boldsymbol{\mu}_i)^\top\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\big)^2\Big] \\
&= \sum_{i=1}^{k}2\omega_i\mathbb{E}\Big[1-\exp\big(-\frac{\|\mathbf{V}_i\|_2^2}{\sigma^2}\big)\Big] \\
&\overset{(d)}{\leq} \sum_{i=1}^{k}2\omega_i\Big(1-\exp\big(-\frac{\mathbb{E}\big[\|\mathbf{V}_i\|_2^2\big]}{\sigma^2}\big)\Big) \\
&\leq \sum_{i=1}^{k}2\omega_i\Big(1-\exp\big(-\frac{\sigma_i^2}{\sigma^2}\big)\Big) \\
&\overset{(e)}{\leq} 2\sum_{i=1}^{k}\omega_i\frac{\sigma_i^2}{\sigma^2}
\end{aligned}$$

In the above, (a) and (b) follow from Jensen's inequality applied to the convex Frobenius-norm-squared function. (c) holds because given the unit-norm vectors $\mathbf{a} = \phi(\boldsymbol{\mu}_i + \mathbf{V}_i)$ and $\mathbf{b} = \phi(\boldsymbol{\mu}_i)$ the following holds

$$\big\|\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top - \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top\big\|_F^2 = \|\mathbf{a}\mathbf{a}^\top - \mathbf{b}\mathbf{b}^\top\|_F^2 = \|\mathbf{a}\|^4 + \|\mathbf{b}\|^4 - 2(\mathbf{a}^\top\mathbf{b})^2 = 2 - 2(\mathbf{a}^\top\mathbf{b})^2.$$

Finally, (d) follows from Jensen's inequality for the concave function $t(z) = 1 - \exp(-z)$, and (e) holds because of the inequality $1 - e^{-t} \leq t$ for every $t \in \mathbb{R}$. Next, we create the following orthogonal basis consisting of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ of the span of the k unit-norm vectors $\phi(\boldsymbol{\mu}_1), \ldots, \phi(\boldsymbol{\mu}_k)$ as follows: We choose $\mathbf{u}_1 = \phi(\boldsymbol{\mu}_1)$, and for every $2 \leq i \leq k$ we construct $\mathbf{u}_i$ as

$$\mathbf{u}_i \;:= \phi(\boldsymbol{\mu}_i) - \sum_{j=1}^{i-1}\langle\phi(\boldsymbol{\mu}_i), \mathbf{u}_j\rangle\mathbf{u}_j.$$

Therefore, we will have:

$$\Big\|\sum_{i=1}^{k}\omega_i\phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top - \sum_{i=1}^{k}\omega_i\mathbf{u}_i\mathbf{u}_i^\top\Big\|_F^2 \;=\; \Big\|\sum_{i=1}^{k}\omega_i\Big(\phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top - \mathbf{u}_i\mathbf{u}_i^\top\Big)\Big\|_F^2$$

$$
\begin{aligned}
&\overset{(f)}{\leq} \sum_{i=1}^{k} \omega_i \Big\| \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top - \mathbf{u}_i\mathbf{u}_i^\top \Big\|_F^2 \\
&\overset{(g)}{=} \sum_{i=1}^{k} \omega_i \Big(1 + \|\mathbf{u}_i\|^4 - 2\big(\mathbf{u}_i^\top \phi(\boldsymbol{\mu}_i)\big)^2\Big) \\
&\leq \sum_{i=1}^{k} \omega_i \Big(2 - 2\big(\mathbf{u}_i^\top \phi(\boldsymbol{\mu}_i)\big)^2\Big) \\
&= \sum_{i=1}^{k} 2\omega_i \Big(1 + \mathbf{u}_i^\top \phi(\boldsymbol{\mu}_i)\Big)\Big(1 - \mathbf{u}_i^\top \phi(\boldsymbol{\mu}_i)\Big) \\
&\overset{(h)}{\leq} \sum_{i=1}^{k}\sum_{j=1}^{i-1} 4\omega_i \exp\Big(\frac{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big).
\end{aligned}
$$

In the above, (f) comes from the application of Jensen's inequality for the convex Frobenius norm-squared function. (g) follows because of the same reason as for item (c). (h) holds because $1 + \mathbf{u}_i^\top \phi(\boldsymbol{\mu}_i) \leq 2$ and

$$
\mathbf{u}_i^\top \phi(\boldsymbol{\mu}_i) = 1 - \sum_{j=1}^{i-1} \langle \phi(\boldsymbol{\mu}_i), \mathbf{u}_j \rangle^2 \geq 1 - \sum_{j=1}^{i-1} \exp\Big(\frac{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big).
$$

Since for every two matrices $A,\ B$ we have $\|A+B\|_F^2 \leq 2\|A\|_F^2 + 2\|B\|_F^2$, we can combine the previous shown inequalities to obtain

$$
\begin{aligned}
\Big\| C_{\mathbf{X}} - \sum_{i=1}^{k} \omega_i \mathbf{u}_i\mathbf{u}_i^\top \Big\|_F^2 \ &\leq\ 2\Big\| C_{\mathbf{X}} - \sum_{i=1}^{k} \omega_i \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top \Big\|_F^2 + 2\Big\| \sum_{i=1}^{k} \omega_i \phi(\boldsymbol{\mu}_i)\phi(\boldsymbol{\mu}_i)^\top - \sum_{i=1}^{k} \omega_i \mathbf{u}_i\mathbf{u}_i^\top \Big\|_F^2 \\
&\leq\ 4\sum_{i=1}^{k} \frac{\omega_i \sigma_i^2}{\sigma^2} + 8\sum_{i=1}^{k}\sum_{j=1}^{i-1} \omega_i \exp\Big(\frac{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big)
\end{aligned}
$$

Since we know that $\|\mathbf{u}_i\|^2\omega_i$ for $i = 1, \ldots, k$ are the eigenvalues of $\sum_{i=1}^{k} \omega_i \mathbf{u}_i\mathbf{u}_i^\top$ where $1 - 2\sum_{j=1}^{i-1} \exp\Big(\frac{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big) \leq \|\mathbf{u}_i\|^2 \leq 1$, then the eigenspectrum stability bound in (Hoffman & Wielandt, 2003) implies that for the top k eigenvalues of $C_{\mathbf{X}}$, denoted by $\lambda_1, \ldots, \lambda_k$, we will have

$$
\sum_{i=1}^{k} \big(\lambda_i - \|\mathbf{u}_i\|^2\omega_i\big)^2 \leq 4\sum_{i=1}^{k} \frac{\omega_i \sigma_i^2}{\sigma^2} + 8\sum_{i=1}^{k}\sum_{j=1}^{i-1} \omega_i \exp\Big(\frac{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big)
$$

Therefore, since $\big(\lambda_i - \|\mathbf{u}_i\|^2\omega_i\big)^2 \leq \big(\lambda_i - \omega_i\big)^2 + 2(1 - \|\mathbf{u}_i\|^2)\omega_i$, we obtain the following which completes the proof:

$$
\sum_{i=1}^{k} \big(\lambda_i - \omega_i\big)^2 \leq 4\sum_{i=1}^{k} \frac{\omega_i \sigma_i^2}{\sigma^2} + 16\sum_{i=1}^{k}\sum_{j=1}^{i-1} \omega_i \exp\Big(\frac{-\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big).
$$

A.1.2. Proof of Theorem 2

To show the theorem, we first follow Theorem 1's proof where we showed that:

$$
\Big\| C_{\mathbf{Y}} - \sum_{i=1}^{k} \gamma_i \phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top \Big\|_F^2 \ \leq\ 2\sum_{i=1}^{k} \frac{\gamma_i \sigma_i^2}{\sigma^2}
$$

Next, we attempt to bound the norm difference between $C_{\mathbf{X}}$ and $\sum_{i=1}^{k} \omega_i \phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top$:

$$
\Big\| C_{\mathbf{X}} - \sum_{i=1}^{k} \omega_i \phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top \Big\|_F^2 = \Big\| \sum_{i=1}^{k} \Big[\omega_i \Big(\mathbb{E}\big[\phi(\boldsymbol{\mu}_i + \mathbf{V}_i)\phi(\boldsymbol{\mu}_i + \mathbf{V}_i)^\top\big] - \phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top \Big) \Big] \Big\|_F^2
$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \sum_{i=1}^{k}\Big[\omega_i\Big\|\mathbb{E}\big[\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top\big]-\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\Big\|_F^2\Big]\\
&= \sum_{i=1}^{k}\Big[\omega_i\Big\|\mathbb{E}\big[\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top-\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\big]\Big\|_F^2\Big]\\
&\overset{(b)}{\leq} \sum_{i=1}^{k}\omega_i\mathbb{E}\Big[\Big\|\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)^\top-\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\Big\|_F^2\Big]\\
&\overset{(c)}{=} \sum_{i=1}^{k}\omega_i\mathbb{E}\Big[2-2\big(\phi(\boldsymbol{\mu}'_i)^\top\phi(\boldsymbol{\mu}_i+\mathbf{V}_i)\big)^2\Big]\\
&= \sum_{i=1}^{k}2\omega_i\mathbb{E}\Big[1-\exp\big(\frac{-\|\mathbf{V}_i+\boldsymbol{\delta}_i\|_2^2}{\sigma^2}\big)\Big]\\
&\overset{(d)}{\leq} \sum_{i=1}^{k}2\omega_i\Big(1-\exp\big(\frac{-\mathbb{E}\big[\|\mathbf{V}_i+\boldsymbol{\delta}_i\|_2^2\big]}{\sigma^2}\big)\Big)\\
&\overset{(e)}{=} \sum_{i=1}^{k}2\omega_i\Big(1-\exp\big(-\frac{\mathbb{E}\big[\|\mathbf{V}_i\|_2^2\big]+\|\boldsymbol{\delta}_i\|_2^2}{\sigma^2}\big)\Big)\\
&\leq \sum_{i=1}^{k}2\omega_i\Big(1-\exp\big(-\frac{\sigma_i^2+\|\boldsymbol{\delta}_i\|_2^2}{\sigma^2}\big)\Big)\\
&\overset{(f)}{\leq} 2\sum_{i=1}^{k}\frac{\omega_i\big(\sigma_i^2+\|\boldsymbol{\delta}_i\|_2^2\big)}{\sigma^2}
\end{aligned}
$$

Note that in the above (a), (b), (c), (d), and (f) hold for the same reason as the same-numbered items hold in the proof of Theorem 1. Also, (e) holds because $\mathbb{E}[\mathbf{V}_i]=\mathbf{0}$.

Then, since for every matrices A, B we have $\|A+B\|_F^2\leq 2\|A\|_F^2+2\|B\|_F^2$, we can combine the above two parts to show:

$$
\begin{aligned}
&\Big\|\Big(C_{\mathbf{X}}-\eta C_{\mathbf{Y}}\Big)-\sum_{i=1}^{k}(\omega_i-\eta\gamma_i)\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\Big\|_F^2\\
=&\Big\|\Big(C_{\mathbf{X}}-\sum_{i=1}^{k}\omega_i\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\Big)-\eta\Big(C_{\mathbf{Y}}-\sum_{i=1}^{k}\gamma_i\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\Big)\Big\|_F^2\\
\leq& 2\Big\|C_{\mathbf{X}}-\sum_{i=1}^{k}\omega_i\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\Big\|_F^2+2\eta^2\Big\|C_{\mathbf{Y}}-\sum_{i=1}^{k}\gamma_i\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top\Big)\Big\|_F^2\\
\leq& 4\sum_{i=1}^{k}\frac{\omega_i\big(\sigma_i^2+\|\boldsymbol{\delta}_i\|_2^2\big)+\eta^2\gamma_i\sigma_i^2}{\sigma^2}\\
=& 4\sum_{i=1}^{k}\frac{\big(\omega_i+\eta^2\gamma_i\big)\sigma_i^2+\omega_i\|\boldsymbol{\delta}_i\|_2^2}{\sigma^2}
\end{aligned}
$$

Next, we create an orthogonal basis consisting of vectors $\mathbf{u}_1,\ldots,\mathbf{u}_t$ of the span of the t unit-norm vectors $\phi(\boldsymbol{\mu}'_1),\ldots,\phi(\boldsymbol{\mu}'_t)$ as follows where for every $1\leq i\leq t$ we construct $\mathbf{u}_i$ as

$$
\mathbf{u}_i\ :=\phi(\boldsymbol{\mu}_i)-\sum_{j=1}^{i-1}\langle\phi(\boldsymbol{\mu}_i),\mathbf{u}_j\rangle\mathbf{u}_j
$$

As a result, we can show:

$$
\Big\|\sum_{i=1}^{k}(\omega_i-\eta\gamma_i)\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top-\sum_{i=1}^{k}(\omega_i-\eta\gamma_i)\mathbf{u}_i\mathbf{u}_i^\top\Big\|_F^2=\Big\|\sum_{i=1}^{k}(\omega_i-\eta\gamma_i)\Big(\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top-\mathbf{u}_i\mathbf{u}_i^\top\Big)\Big\|_F^2
$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{k}(1+\eta)(\omega_i+\eta\gamma_i)\Big\|\phi(\boldsymbol{\mu}'_i)\phi(\boldsymbol{\mu}'_i)^\top - \mathbf{u}_i\mathbf{u}_i^\top\Big\|_F^2 \\
&= \sum_{i=1}^{k}(1+\eta)(\omega_i+\eta\gamma_i)\Big(1+\|\mathbf{u}_i\|^4 - 2\big(\mathbf{u}_i^\top\phi(\boldsymbol{\mu}'_i)\big)^2\Big) \\
&\leq \sum_{i=1}^{k}\sum_{j=1}^{i-1}4(1+\eta)(\omega_i+\eta\gamma_i)\exp\Big(\frac{-\|\boldsymbol{\mu}'_i-\boldsymbol{\mu}'_j\|_2^2}{\sigma^2}\Big).
\end{aligned}
$$

Therefore, we can combine the above results to show:

$$
\begin{aligned}
&\Big\|\big(C_{\mathbf{X}}-\eta C_{\mathbf{Y}}\big)-\sum_{i=1}^{k}(\omega_i-\eta\gamma_i)\mathbf{u}_i\mathbf{u}_i^\top\Big\|_F^2 \\
\leq& 8\sum_{i=1}^{k}\frac{\big(\omega_i+\eta^2\gamma_i\big)\sigma_i^2+\omega_i\|\boldsymbol{\delta}_i\|_2^2}{\sigma^2} \\
&+8(1+\eta)\sum_{i=1}^{k}\sum_{j=1}^{i-1}(\omega_i+\eta\gamma_i)\exp\Big(\frac{-\|\boldsymbol{\mu}'_i-\boldsymbol{\mu}'_j\|_2^2}{\sigma^2}\Big).
\end{aligned}
$$

Since we know that $\|\mathbf{u}_i\|^2(\omega_i-\eta\gamma_i)$ for $i=1,\ldots,t$ are the eigenvalues of $\sum_{i=1}^{k}(\omega_i-\eta\gamma_i)\mathbf{u}_i\mathbf{u}_i^\top$ where $1-2\sum_{j=1}^{i-1}\exp\Big(\frac{-\|\boldsymbol{\mu}'_i-\boldsymbol{\mu}_j\|_2^2}{\sigma^2}\Big)\leq\|\mathbf{u}_i\|^2\leq 1$, then the eigenspectrum stability bound in (Hoffman & Wielandt, 2003) shows that for the top k eigenvalues of $C_{\mathbf{X}}-\eta C_{\mathbf{Y}}$, denoted by $\lambda_1\geq\ldots\geq\lambda_k$, we will have

$$
\begin{aligned}
&\sum_{i=1}^{k}\big(\lambda_i-\|\mathbf{u}_i\|^2(\omega_i-\eta\gamma_i)\big)^2 \\
\leq& 8\sum_{i=1}^{k}\frac{\big(\omega_i+\eta^2\gamma_i\big)\sigma_i^2+\omega_i\|\boldsymbol{\delta}_i\|_2^2}{\sigma^2}+8(1+\eta)\sum_{i=1}^{k}\sum_{j=1}^{i-1}(\omega_i+\eta\gamma_i)\exp\Big(\frac{-\|\boldsymbol{\mu}'_i-\boldsymbol{\mu}'_j\|_2^2}{\sigma^2}\Big).
\end{aligned}
$$

As a consequence, since $\big(\lambda_i-\|\mathbf{u}_i\|^2(\omega_i-\eta\gamma_i)\big)^2\leq\big(\lambda_i-(\omega_i-\eta\gamma_i)\big)^2+2(1-\|\mathbf{u}_i\|^2)\max\{\omega_i-\eta\gamma_i,0\}$ and $\mathrm{ReLU}(z)=\max\{z,0\}$ is a 1-Lipschitz function, we obtain the following which finishes the proof:

$$
\begin{aligned}
&\sum_{i=1}^{k}\big(\max\{\lambda_i,0\}-\max\{\omega_i-\eta\gamma_i,0\}\big)^2 \\
\leq& 8\sum_{i=1}^{k}\frac{\big(\omega_i+\eta^2\gamma_i\big)\sigma_i^2+\omega_i\|\boldsymbol{\delta}_i\|_2^2}{\sigma^2}+16(1+\eta)\sum_{i=1}^{k}\sum_{j=1}^{i-1}(\omega_i+\eta\gamma_i)\exp\Big(\frac{-\|\boldsymbol{\mu}'_i-\boldsymbol{\mu}'_j\|_2^2}{\sigma^2}\Big)
\end{aligned}
$$

A.1.3. Proof of Theorem 3

We note that given the empirical kernel feature matrices $\Phi_{\mathbf{X}}\in\mathbb{R}^{n\times s}$ and $\Phi_{\mathbf{X}}\in\mathbb{R}^{m\times s}$, we can write

$$
\widehat{C}_{\mathbf{X}}=\frac{1}{n}\Phi_{\mathbf{X}}^\top\Phi_{\mathbf{X}},\quad \widehat{C}_{\mathbf{Y}}=\frac{1}{m}\Phi_{\mathbf{Y}}^\top\Phi_{\mathbf{Y}}.
$$

Therefore, defining $\tilde{\Phi}_{\mathbf{X}}=\frac{1}{\sqrt{n}}\Phi_{\mathbf{X}}$ and $\tilde{\Phi}_{\mathbf{Y}}=\frac{1}{\sqrt{m}}\Phi_{\mathbf{Y}}$, we can rewrite the definition of the η-differential kernel covariance matrix as

$$
\begin{aligned}
\widehat{C}_{\mathbf{X}}-\eta\widehat{C}_{\mathbf{Y}}&=\tilde{\Phi}_{\mathbf{X}}^\top\tilde{\Phi}_{\mathbf{X}}-\eta\tilde{\Phi}_{\mathbf{Y}}^\top\tilde{\Phi}_{\mathbf{Y}} \\
&=\begin{bmatrix}\tilde{\Phi}_{\mathbf{X}}\\ \sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}}\end{bmatrix}^\top\begin{bmatrix}\tilde{\Phi}_{\mathbf{X}}\\ -\sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}}\end{bmatrix}.
\end{aligned}
$$

Defining $A = \begin{bmatrix} \tilde{\Phi}_{\mathbf{X}} \\ \sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}} \end{bmatrix}$ and $B = \begin{bmatrix} \tilde{\Phi}_{\mathbf{X}} \\ -\sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}} \end{bmatrix}$, we use the property that $A^\top B$ and $BA^\top$ share the same non-zero eigenvalues, because if for $\lambda \neq 0$ and $\mathbf{v}$ we have $A^\top B\mathbf{v} = \lambda\mathbf{v}$, then for $\mathbf{u} = B\mathbf{v}$ we have $BA^\top\mathbf{u} = \lambda\mathbf{u}$. Therefore, the non-zero eigenvalues of the η-differential kernel covariance matrix $\Lambda_{\mathbf{X}|\eta\mathbf{Y}} = \widehat{C}_{\mathbf{X}} - \eta\widehat{C}_{\mathbf{Y}}$ will be the same as the non-zero eigenevalues of

$$\begin{aligned}
\begin{bmatrix} \tilde{\Phi}_{\mathbf{X}} \\ -\sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}} \end{bmatrix} \begin{bmatrix} \tilde{\Phi}_{\mathbf{X}} \\ \sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}} \end{bmatrix}^\top &= \begin{bmatrix} \tilde{\Phi}_{\mathbf{X}}\tilde{\Phi}_{\mathbf{X}}^\top & \sqrt{\eta}\tilde{\Phi}_{\mathbf{X}}\tilde{\Phi}_{\mathbf{Y}}^\top \\ -\sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}}\tilde{\Phi}_{\mathbf{X}}^\top & -\eta\tilde{\Phi}_{\mathbf{Y}}\tilde{\Phi}_{\mathbf{Y}}^\top \end{bmatrix} \\
&= \begin{bmatrix} K_{\mathbf{XX}} & \sqrt{\eta}\,K_{\mathbf{XY}} \\ -\sqrt{\eta}\,K_{\mathbf{XY}}^\top & -\eta\,K_{\mathbf{YY}} \end{bmatrix} \\
&= K_{\mathbf{X}|\eta\mathbf{Y}}.
\end{aligned}$$

In addition, given every eigenvector $\mathbf{v}$ of $K_{\mathbf{X}|\eta\mathbf{Y}}$, the vector $\mathbf{u} = A^\top\mathbf{v} = \begin{bmatrix} \tilde{\Phi}_{\mathbf{X}} \\ \sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}} \end{bmatrix}^\top \mathbf{v}$ will be an eigenvector of $\Lambda_{\mathbf{X}|\eta\mathbf{Y}} = \widehat{C}_{\mathbf{X}} - \eta\widehat{C}_{\mathbf{Y}}$, which will be

$$\begin{bmatrix} \tilde{\Phi}_{\mathbf{X}} \\ \sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}} \end{bmatrix}^\top \mathbf{v} = \sum_{i=1}^{n} v_i\phi(\mathbf{x}_i) + \sum_{j=1}^{m} \sqrt{\eta}v_{n+j}\phi(\mathbf{y}_j).$$

Therefore, the proof is complete.

A.1.4. Proof of Theorem 4

First, we note that $K_{\mathbf{X},\eta\mathbf{Y}}$ is a symmetric PSD matrix, because defining $A = \begin{bmatrix} \tilde{\Phi}_{\mathbf{X}} \\ \sqrt{\eta}\tilde{\Phi}_{\mathbf{Y}} \end{bmatrix}$ where $\tilde{\Phi}_{\mathbf{X}} = \frac{1}{\sqrt{n}}\Phi_{\mathbf{X}}$ and $\tilde{\Phi}_{\mathbf{Y}} = \frac{1}{\sqrt{m}}\Phi_{\mathbf{Y}}$ we will have $K_{\mathbf{X},\eta\mathbf{Y}} = AA^\top$. Therefore, applying the Cholesky decomposition, we can find a $V \in \mathbb{R}^{(m+n)\times(m+n)}$ such that $K_{\mathbf{X},\eta\mathbf{Y}} = V^\top V$.

Next, we note the following identity given $D = \mathrm{diag}\{[\underbrace{+1,\ldots,+1}_{\text{n times}},\underbrace{-1,\ldots,-1}_{\text{m times}}]\}$:

$$\begin{aligned}
K_{\mathbf{X}|\eta\mathbf{Y}} &= \begin{bmatrix} K_{\mathbf{XX}} & \sqrt{\eta}\,K_{\mathbf{XY}} \\ -\sqrt{\eta}\,K_{\mathbf{XY}}^\top & -\eta\,K_{\mathbf{YY}} \end{bmatrix} \\
&= D\begin{bmatrix} K_{\mathbf{XX}} & \sqrt{\eta}\,K_{\mathbf{XY}} \\ \sqrt{\eta}\,K_{\mathbf{XY}}^\top & \eta\,K_{\mathbf{YY}} \end{bmatrix} \\
&= DK_{\mathbf{X},\eta\mathbf{Y}} \\
&= DV^\top V.
\end{aligned}$$

However, we observe that based on the same argument in Theorem 2's proof, $DV^\top V$ and $VDV^\top$ have the same non-zero eigenvalues. Therefore, the symmetric matrix $VDV^\top$ and $K_{\mathbf{X}|\eta\mathbf{Y}}$ share the same non-zero eigenvalues.

A.2. Experimental Results

A.2.1. Applications of KEN Score

Missing mode detection. To enable missing mode detection, we can select a large enough η for $K_{\mathbf{X}|\eta\mathbf{Y}}$. For example, according to Figure 6, the modes "Microphone", "Round hat", and "Black uniform hat" are found missing in LDM by its training set and other generative models.

Specific novel mode generation. The qualitative analysis can reveal most related samples of a novel mode. Therefore, we can retrieve the latent z of these novel samples to fit a Gaussian. Then, we sample from this Gaussian to obtain new samples

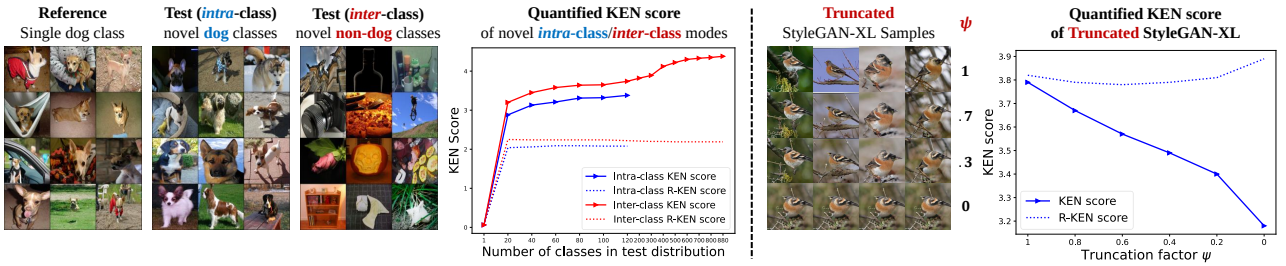

Figure 5. Quantified KEN score in real and generated distributions. **Left:** KEN score in ImageNet-1K. *Intra*-class means similarity in taxonomy (e.g. Dogs with different breeds). **Right:** KEN score in truncated StyleGAN-XL. ψ is truncation factor. $\psi = 1$ reduces to normal StyleGAN-XL. *"R-KEN"* means switching test and reference distributions. Inception-V3 embedding is used.

in the same novel mode. We put an example of specifically generating more FFHQ "kids" with StyleGAN-XL in Figure 7.

Benchmarking mode novelty. For a group of generative models with the same training set. We can evaluate the mode novelty between them. The average novelty of a generative model to others can be used for benchmarking. Table 1 shows mode novelty between generative models trained on FFHQ. We observe that InsGen has the highest average novelty and VDVAE has the lowest average novelty in this group.

Benchmarking fitness. If we use generative models and their training sets as testing and reference distribution, our proposed KEN can be recognized as a divergence measurement. When two distributions are identical, their KEN evaluation will be 0. In Table 2, we observe KEN behave similarly with FID in ImageNet and CIFAR-10, except for GGAN, DCGAN, and WGAN in CIFAR-10.

A.2.2. Extra Quantified Analysis of KEN Score

Distinct modes contain richer novelty. To define similar modes, we extract 120 dog classes from ImageNet-1K. The remaining 880 classes are dog-excluded and represent distinct modes. We select a single dog class as the reference, other dog classes as novel *intra*-class modes, and 880 dog-excluded classes as novel *inter*-class modes. Figure 5 shows that adding novel modes to test distribution increases mode novelty. Meanwhile, the line chart in Figure 5 indicates *inter*-class modes contain richer novelty than *intra*-class modes since the red *inter*-class line is higher. The reversed novelty lines remain flat, illustrating the asymmetric property.

Truncation trick decreases mode novelty. Truncation trick (Marchesi, 2017; Brock et al., 2018) is a procedure sampling latent z from a truncated normal to trade-off diversity for high-fidelity generated images. We observe this trick also reduces the KEN score of generative model in Figure 5.

A.2.3. Extra Experimental Results

Additional real dataset results. Figure 8 shows detected novel modes between more real datasets. The novel modes of CelebA to FFHQ relate to the background of celebrities. For dog subsets of ImageNet and AFHQ, ImageNet-dogs seems to be novel in the dog breeds, while AFHQ-dogs seem to have more young dogs than ImageNet-dogs. Figure 9 shows novel modes of all possible pairs of generative models in Figure 4.

Novel modes detection with different embedding. Figure 10, 11, 12 shows the detected top-9 novel modes of the AFHQ dataset with respect to the ImageNet-dogs dataset with Inception-V3, DINOv2, and CLIP embedding, respectively. The choice of embedding affect the detected novel modes and their rankings by the proposed KEN method.

Generative models' KEN scores with different embedding. Table 3, 4, and 5 shows KEN scores between generative models trained on the FFHQ dataset with Inception-V3, DINOv2, and CLIP embedding. We observed the rankings of the average KEN score are consistent with the same embedding but different choices of bandwidth parameter σ. However, the average KEN score ranking of generative models evaluated by the Inception-V3 embedding is different from rankings evaluated with DINOv2 and CLIP embedding.

Figure 6. Missing modes of LDM in FFHQ. Training set and other generative models both capture similar missing modes of LDM with large $\eta = 10$ in $K_{X|Y}$. Inception-V3 embedding is used.

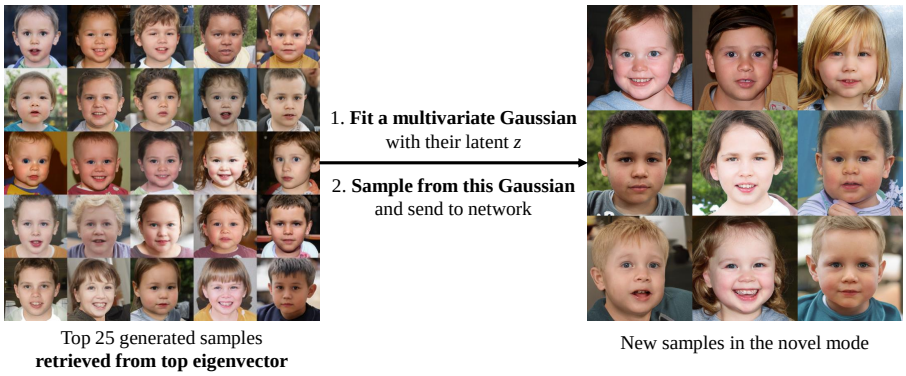

Figure 7. Generating new samples in a specific novel mode by fitting a Gaussian with samples from qualitative analysis. Inception-V3 embedding is used.

Table 2. Benchmarking fitness of generative models. More powerful models tend to have lower KEN to the training set. $\sigma = 10$. Inception-V3 embedding is used.

Dataset	Model	IS	FID	Precision	Recall	Density	Coverage	KEN
CIFAR10	GGAN (Lim & Ye, 2017)	6.51	40.22	0.56	0.30	0.42	0.31	3.03
	DCGAN (Radford et al., 2015)	5.76	51.98	0.62	0.16	0.56	0.25	3.02
	WGAN-WC (Arjovsky et al., 2017)	3.99	95.69	0.53	0.04	0.40	0.11	3.01
	WGAN-GP (Gulrajani et al., 2017)	7.04	26.42	0.62	0.56	0.55	0.46	2.99
	ACGAN (Odena et al., 2017)	7.02	35.42	0.60	0.23	0.50	0.32	2.99
	LSGAN (Mao et al., 2017)	7.13	31.31	0.61	0.41	0.50	0.42	2.97
	LOGAN (Wu et al., 2019)	7.95	17.86	0.64	0.64	0.60	0.56	2.90
	SAGAN (Zhang et al., 2019)	8.67	9.58	0.69	0.63	0.72	0.72	2.70
	SNGAN (Miyato et al., 2018)	8.77	8.50	0.71	0.62	0.79	0.75	2.65
	BigGAN (Brock et al., 2018)	9.14	6.80	0.71	0.61	0.86	0.80	2.59
	ContraGAN (Kang & Park, 2021)	9.40	6.55	0.73	0.61	0.87	0.81	2.57
	StyleGAN2-ADA (Karras et al., 2020)	10.14	3.61	0.73	0.67	0.98	0.89	2.50
ImageNet 128^2	SAGAN (Zhang et al., 2019)	14.47	64.04	0.33	0.54	0.16	0.14	3.46
	StyleGAN2-SPD (Karras et al., 2019)	21.08	35.27	0.50	0.62	0.37	0.33	3.17
	StyleGAN3-t-SPD (Karras et al., 2021)	20.90	33.69	0.52	0.61	0.38	0.32	3.13
	SNGAN (Miyato et al., 2018)	32.28	28.66	0.54	0.67	0.42	0.41	3.07
	ContraGAN (Kang & Park, 2021)	25.19	28.33	0.67	0.53	0.64	0.34	2.91
	ReACGAN (Kang et al., 2021)	52.95	18.19	0.76	0.40	0.88	0.49	2.67
	BigGAN-2048 (Brock et al., 2018)	104.57	11.92	0.74	0.40	0.98	0.75	2.56
	StyleGAN-XL (Sauer et al., 2022)	225.16	2.71	0.80	0.63	1.12	0.93	2.42

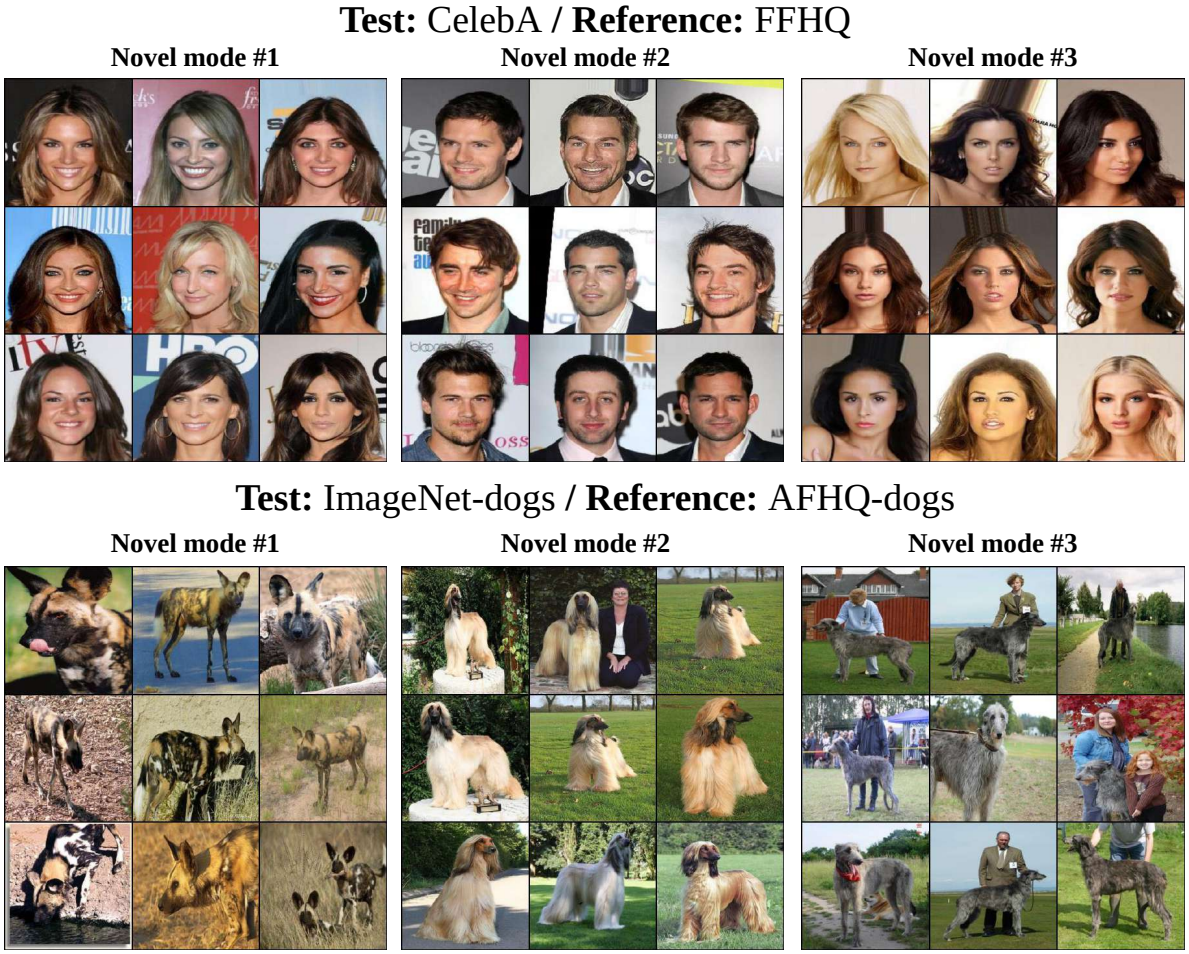

Figure 8. Novel modes between real datasets visualized with top-3-ranked eigenvectors. Extra samples of Figure 3. Inception-V3 embedding is used.

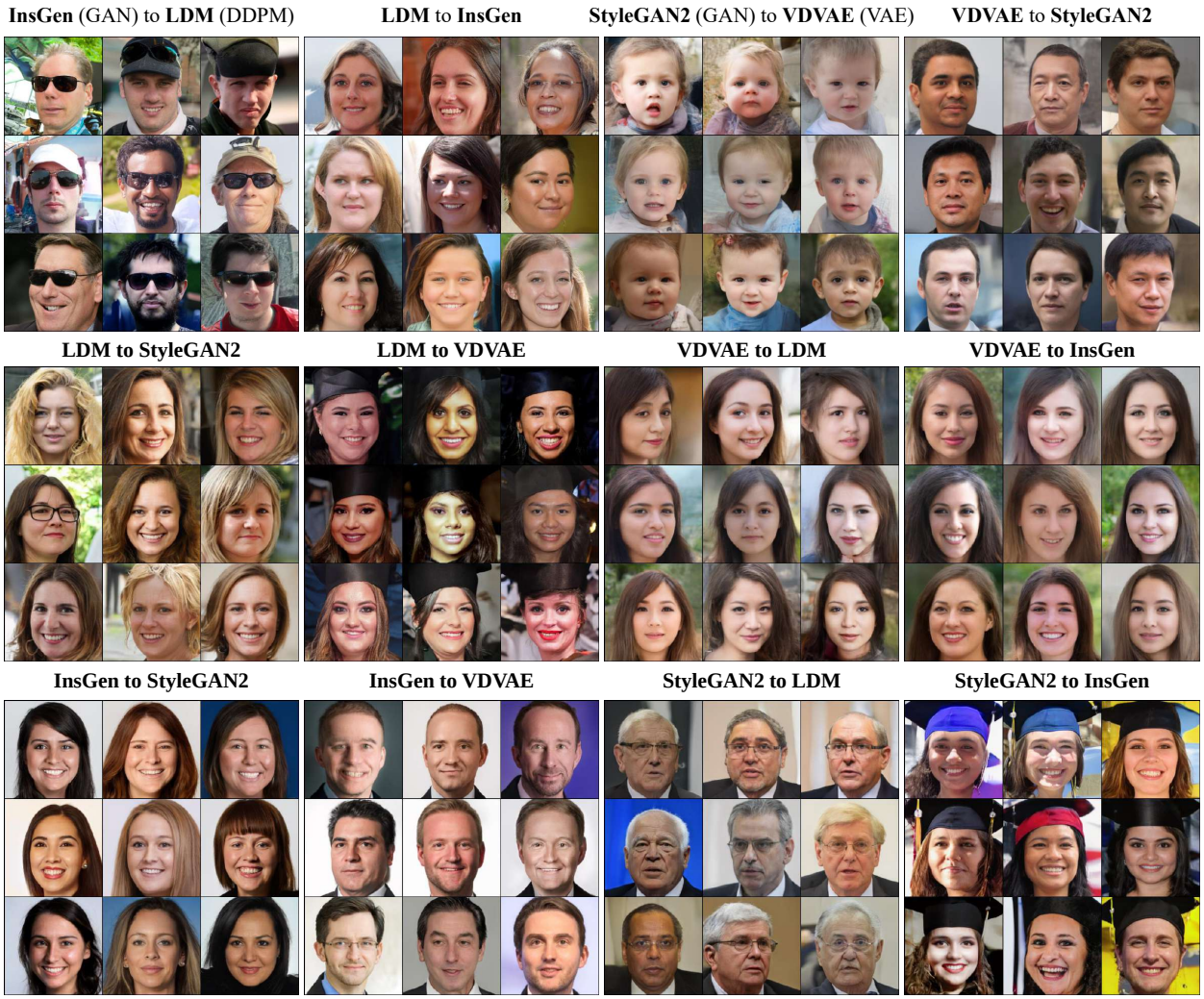

Figure 9. Novel modes between FFHQ-trained generative models in various architecture with the top-ranked eigenvector. Extra samples of Figure 4. Inception-V3 embedding is used.

Table 3. FFHQ-trained generative models' pairwise KEN score. Inception-V3 embedding is used.

Bandwidth σ	Generative Models (Test Models)	Reference Models					Avg. KEN
		InsGen	StyleGAN-XL	StyleGAN2	LDM	VDVAE	
10	InsGen (Yang et al., 2021)	-	3.56	3.55	3.64	4.27	3.76
	StyleGAN-XL (Sauer et al., 2022)	3.57	-	3.61	3.61	4.21	3.75
	StyleGAN2 (Karras et al., 2019)	3.46	3.54	-	3.60	4.08	3.67
	LDM (Rombach et al., 2021)	3.45	3.45	3.49	-	3.97	3.59
	VDVAE (Child, 2020)	3.26	3.24	3.19	3.17	-	3.22
15	InsGen (Yang et al., 2021)	-	1.17	1.18	1.26	1.87	1.37
	StyleGAN-XL (Sauer et al., 2022)	1.16	-	1.19	1.24	1.83	1.36
	StyleGAN2 (Karras et al., 2019)	1.12	1.18	-	1.26	1.76	1.33
	LDM (Rombach et al., 2021)	1.09	1.08	1.14	-	1.59	1.23
	VDVAE (Child, 2020)	0.96	0.95	0.94	0.91	-	0.94

Figure 10. Top 9 novel modes of the AFHQ dataset w.r.t. the ImageNet-dogs dataset. Inception embedding is used.

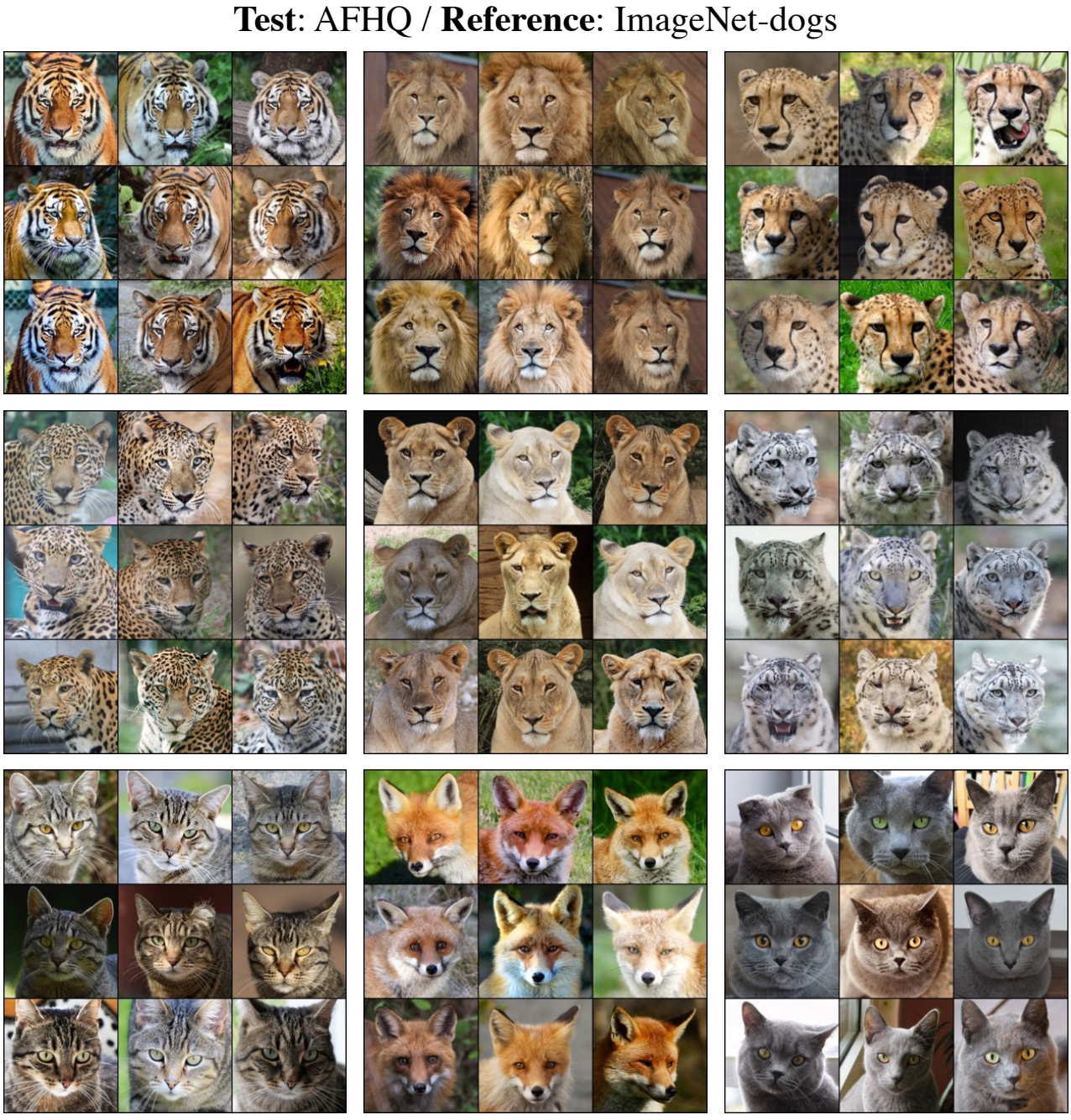

Figure 11. Top 9 novel modes of the AFHQ dataset w.r.t. the ImageNet-dogs dataset. DINOv2 embedding is used.

Figure 12. Top 9 novel modes of the AFHQ dataset w.r.t. the ImageNet-dogs dataset. CLIP embedding is used.

Table 4. FFHQ-trained generative models' pairwise KEN score. DINOv2 embedding is used.

Bandwidth σ	Generative Models (Test Models)	Reference Models					Avg. KEN
		StyleGAN-XL	LDM	InsGen	StyleGAN2	VDVAE	
30	StyleGAN-XL (Sauer et al., 2022)	-	4.35	4.53	4.59	4.75	4.56
	LDM (Rombach et al., 2021)	4.25	-	4.28	4.37	4.52	4.36
	InsGen (Yang et al., 2021)	4.39	4.24	-	3.92	4.70	4.31
	StyleGAN2 (Karras et al., 2019)	4.38	4.27	3.87	-	4.64	4.29
	VDVAE (Child, 2020)	4.17	4.02	4.22	4.24	-	4.16
50	StyleGAN-XL (Sauer et al., 2022)	-	1.48	1.60	1.65	1.84	1.64
	LDM (Rombach et al., 2021)	1.41	-	1.46	1.53	1.71	1.53
	InsGen (Yang et al., 2021)	1.50	1.44	-	1.25	1.83	1.51
	StyleGAN2 (Karras et al., 2019)	1.50	1.47	1.21	-	1.79	1.49
	VDVAE (Child, 2020)	1.42	1.34	1.47	1.49	-	1.43

Table 5. FFHQ-trained generative models' pairwise KEN score. CLIP embedding is used.

Bandwidth σ	Generative Models (Test Models)	Reference Models					Avg. KEN
		StyleGAN-XL	LDM	InsGen	StyleGAN2	VDVAE	
5	StyleGAN-XL (Sauer et al., 2022)	-	4.11	3.82	3.92	4.26	4.03
	LDM (Rombach et al., 2021)	4.03	-	3.85	3.80	4.30	4.00
	InsGen (Yang et al., 2021)	3.80	3.88	-	3.64	4.26	3.90
	StyleGAN2 (Karras et al., 2019)	3.84	3.77	3.57	-	4.10	3.82
	VDVAE (Child, 2020)	3.49	3.58	3.49	3.46	-	3.51
10	StyleGAN-XL (Sauer et al., 2022)	-	0.90	0.77	0.82	1.10	0.90
	LDM (Rombach et al., 2021)	0.87	-	0.79	0.79	1.11	0.89
	InsGen (Yang et al., 2021)	0.79	0.82	-	0.72	1.12	0.86
	StyleGAN2 (Karras et al., 2019)	0.80	0.78	0.69	-	1.05	0.83
	VDVAE (Child, 2020)	0.73	0.75	0.73	0.73	-	0.74