IMAGE QUALITY EMERGES FROM THE GAUSSIANITY OF DIFFUSION INVERSION

Anonymous authors
Paper under double-blind review

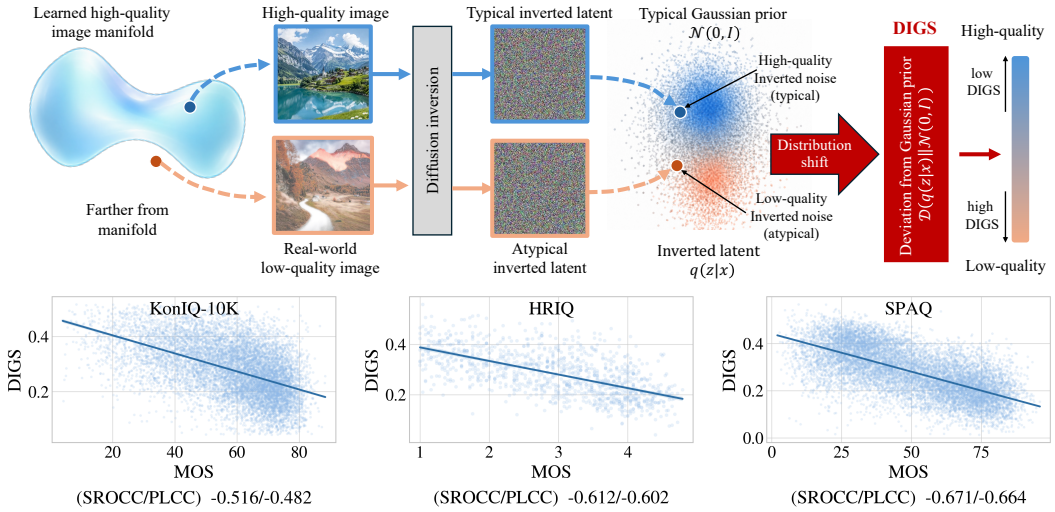


Figure 1: (Top) Diffusion inversion reveals image–prior compatibility. High-quality images tend to invert to more typical Gaussian latents, whereas real-world low-quality images yield more atypical ones. (Bottom) DIGS correlates with Human Quality Judgments and exhibits a clear monotonic correlation with MOS.

ABSTRACT

Diffusion models learn to generate natural images by transforming Gaussian noise into samples from a complex image distribution. While previous studies mainly exploit this forward generation process, we ask whether the inverse trajectory can reveal information about the observed image itself. We discover that diffusion inversion provides a measurable signal of observation–prior compatibility: high-quality images that better agree with the learned image distribution tend to recover terminal latents with Gaussian-like statistics, whereas low-quality images induce increasingly atypical latent distributions. Based on this finding, we introduce Diffusion Inversion Gaussianity Score (DIGS), a training-free and quality-label-free measure that evaluates image quality by quantifying the deviation of inversion latents from the intrinsic Gaussian prior. Beyond marginal statistics, we further characterize channel-wise and spatial deviations to obtain DIGS*. Extensive experiments demonstrate that diffusion inversion statistics correlate strongly with human perceptual judgments across authentic and synthetic distortions, while requiring no quality annotations, reference images, or additional training. Our results reveal diffusion inversion as a new measurement space for probing how observations interact with learned generative priors.

1 INTRODUCTION

Diffusion models (Sohl-Dickstein et al., 2015; Rombach et al., 2022a; Song et al., 2021a; Ho et al., 2020; Song et al., 2021b) have emerged as powerful generative models by learning a transformation between a simple Gaussian prior and the complex distribution of natural images. Most existing

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

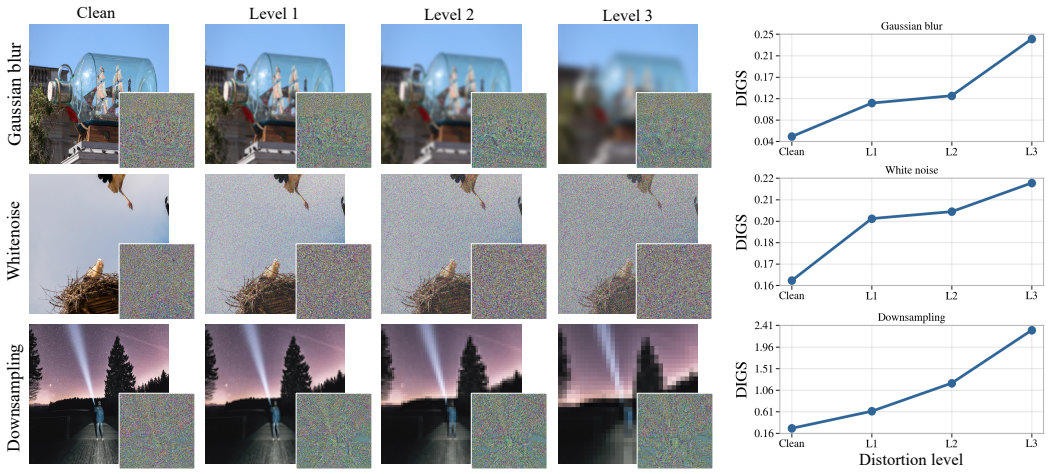


Figure 2: Illustration of DIGS monotonicity for different degradations.

studies (Dhariwal & Nichol, 2021; Song et al., 2021a) focus on the forward generation process, investigating how meaningful images are synthesized from random noise. However, the inverse direction remains largely unexplored:

Can the Gaussianity of diffusion inversion noise reveal the compatibility between an image and a learned generative prior?

Diffusion inversion (Wallace et al., 2023; Karras et al., 2022) has been widely studied for reconstruction (Zeng et al., 2025; Garibi et al., 2024), editing (Mokady et al., 2023; Dong et al., 2023), and personalization (Xing et al., 2025), where the recovered latent is mainly used as an intermediate representation. Recent studies (Staniszewski et al., 2026) investigated the relationship between inversion noise and image representations, revealing that smooth image areas have less diverse noise. However, whether such statistical deviations carry meaningful information about the compatibility between an image and the learned generative prior remains unexplored. In this work, we investigate a different perspective: whether the statistical properties of inverted noise itself provide information about how an image is explained by a pretrained generative model. Rather than treating inversion as only a latent recovery procedure, we study it as a measurement process for probing image–prior compatibility.

We discover that the Gaussianity of diffusion inversion noise provides a measurable signature of this compatibility. Images that better align with the learned image distribution tend to produce inversion noise with statistics closer to the intrinsic Gaussian prior, whereas degraded images induce increasingly atypical latent distributions, as shown in Figure 1. This suggests that diffusion inversion preserves not only image content information, but also statistical traces of how an observation agrees with the learned generative prior.

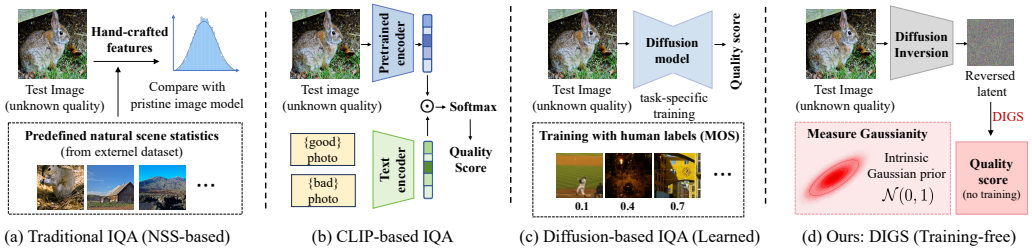
Based on this observation, we hypothesize that:

$$\textit{Image–prior mismatch} \iff \textit{Inversion Gaussianity mismatch}$$

Importantly, this does not assume that generative-prior compatibility is identical to human perception. Instead, we investigate whether inversion Gaussianity provides a consistent statistical signal associated with perceptual quality across diverse image distributions.

We study this phenomenon through no-reference image quality assessment (IQA). Unlike existing approaches (Mittal et al., 2012b; Golestaneh et al., 2022; Ke et al., 2021) that learn quality predictors from human annotations or rely on external natural-image statistics, we ask whether a pretrained diffusion model already contains an intrinsic statistical reference that can be directly accessed through inversion. In this view, IQA serves as a testbed for evaluating whether diffusion inversion can reveal meaningful relationships between observations and generative priors.

Based on this discovery, we introduce **Diffusion Inversion Gaussianity Score (DIGS)**, a training-free and quality-label-free measure that quantifies image–prior compatibility through the deviation of inversion noise from the Gaussian prior. We further develop DIGS* by incorporating complemen-



(a) Traditional IQA (NSS-based) (b) CLIP-based IQA (c) Diffusion-based IQA (Learned) (d) Ours: DIGS (Training-free)

Figure 3: Comparisons of different no-reference IQA.

tary channel-wise and spatial statistics to capture richer structures beyond marginal Gaussianity. Extensive experiments demonstrate that inversion Gaussianity consistently correlates with human perceptual judgments in Figure 1, and remains effective under different inversion settings and diffusion backbones. Across different distortion types, DIGS consistently captures progressive deviation from the diffusion prior, as shown in Figure 2.

Our contributions are summarized as follows:

- We reveal a previously unexplored property of diffusion inversion: the Gaussianity of recovered noise provides a statistical signature of image–prior compatibility.
- We propose DIGS, a **training-free, quality-label-free, no-reference** quality signal that uses diffusion inversion noise as a probe of generative-prior compatibility.
- We develop DIGS* by incorporating channel-wise and spatial statistics, showing that compatibility information exists beyond marginal Gaussianity.
- Through extensive experiments, we demonstrate that diffusion inversion provides a new perspective for understanding the relationship between observations and learned generative priors.

2 Related Work

No-reference image quality assessment (NR-IQA) aims to estimate perceptual quality without access to pristine reference images. Figure 3 and Table 1 summarizes the NR-IQA paradigms and highlights the difference between existing approaches and our training-free measurement perspective.

Natural-scene statistics. Natural-scene-statistics (NSS) methods treat distortion as a disruption of regularities in natural imagery. BRISQUE describes locally normalized luminance statistics and learns their mapping to human opinion scores (Mittal et al., 2012a). NIQE removes this quality-label training by fitting a multivariate Gaussian to features from pristine natural images and measuring the distance of a test image from that model (Mittal et al., 2012b). IL-NIQE expands this idea with richer cues and local modeling (Zhang et al., 2015), whereas PIQE (introduced as PIQUE) discards the external natural-image model and scores perceptually active blocks using fixed rules (Venkatanath et al., 2015). This statistical perspective has also been extended through alternative spatial, frequency-domain, and multi-scale descriptors (Moorthy & Bovik, 2010; 2011; Saad et al., 2012; Ye et al., 2012; Xue et al., 2014; Xu et al., 2016). Across these variants, statistical BIQA has become increasingly independent of annotated distortions, but remains dependent on its underlying assumptions: the descriptors, pristine corpus, region selection, and thresholds still determine what the metric can detect.

Frozen foundation-model readouts. Foundation models replace handcrafted naturalness with knowledge acquired through large-scale pretraining. Related foundation-model approaches derive quality signals from pretrained visual, semantic, or multimodal representations (Kang et al., 2014; Zhang et al., 2018; Pan et al., 2022; Zhang et al., 2023). Built on CLIP's image–language representation (Radford et al., 2021), CLIP-IQA (Wang et al., 2023)scores an image against paired quality antonyms while keeping the base model frozen and requiring no IQA-specific fitting. Its CLIP-IQA+ variant instead learns the prompts from quality annotations, illustrating that a frozen prior and a training-free readout are distinct choices. The language-defined axis also mixes low-level appearance with semantic and aesthetic information. CLIP-IQA therefore demonstrates direct scoring from a general prior, but it does not measure how an image relates to a generative noise distribution.

Table 1: Comparison of DIGS with existing IQA paradigms.

Method category	Traditional IQA (NSS-based)	CLIP-based IQA)	Diffusion-based IQA (Learned)	**Ours: DIGS (Training-free)**
Main idea	Hand-crafted image statistics	Vision-language priors	Train or adapt diffusion model	**Inversion latent atypicality**
Requires quality labels	×	×	✓	×
Requires pristine reference set	✓ (for statistics fitting)	×	×	×
Requires training	×	×	✓	×
Reference distribution	External natural image model	CLIP semantic embedding space	Learned diffusion representations	**Intrinsic Gaussian prior**

Diffusion priors for learned IQA. Recent BIQA methods exploit diffusion models as richer task representations. DP-IQA extracts multi-level U-Net features at a fixed timestep and trains adapters, a quality decoder, and a regressor (Fu et al., 2024); and LGDM guides latent denoising to obtain multi-scale, multi-timestep hyperfeatures before fitting a quality regressor (Saini et al., 2025). A related direction makes denoising itself quality-oriented: PFD-IQA learns a feature-space diffusion process from the pseudo-features and labels of an IQA teacher (Li et al., 2025), while DiffV2IQA trains a restoration model from distorted–reference pairs and uses its outputs in a supervised quality network (Wang et al., 2025). Further learned IQA methods have similarly explored denoising features, generative priors, and restoration trajectories as perceptual quality signals (Sang et al., 2025; Babnik et al., 2024). Despite their differing designs, these diffusion-based methods generally require quality labels, teachers, paired data, or other forms of task-specific learning. They thus address how to *learn* an IQA predictor around a diffusion prior, leaving open whether the diffusion representation itself admits a fixed statistical readout.

Taken together, existing no-reference IQA paradigms rely on external natural-image statistics, indirect quality cues from pretrained representations, or task-specific predictors learned with quality supervision. In contrast, DIGS uses the intrinsic Gaussian prior of a frozen diffusion model as its reference and assesses quality through the statistical atypicality of the terminal inversion latent. It requires neither quality labels, a pristine reference set, nor task-specific training, providing a fully no-reference and training-free quality measure.

3 Diffusion Inversion Gaussianity Score

3.1 Motivation

Diffusion models establish a remarkable correspondence between a simple Gaussian prior and the complex distribution of natural images. Given a diffusion model trained predominantly on high-quality (HQ) natural images, the diffusion model learns a mapping Gaussian noise $p_{\text{noise}}(\boldsymbol{z})$ to the distribution of high-quality natural images $p_{\text{HQ}}(\boldsymbol{x})$:

$$p_{\text{noise}}(\boldsymbol{z}) \longrightarrow p_{\text{HQ}}(\boldsymbol{x}), \tag{1}$$

This observation raises an intriguing question: *if Gaussian noise is naturally associated with high-quality images through the learned diffusion process, what noise will be recovered when we invert a real-world low-quality image?*

For an HQ image $\boldsymbol{x}^{\text{HQ}}$ that is well aligned with the learned image distribution, diffusion inversion approximately reverses the generative trajectory to get $\hat{\boldsymbol{z}}_T^{\text{HQ}}$. The recovered latent is expected to remain in a typical region of the Gaussian prior, as shown in Figure 1,

$$q\left(\hat{\boldsymbol{z}}_T^{\text{HQ}} \mid \boldsymbol{x}^{\text{HQ}}\right) \approx \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{2}$$

Real-world low-quality (LQ) images, however, may exhibit unknown degradations that make their statistics deviate from the HQ distribution learned by the diffusion model. For an LQ observation $\boldsymbol{x}^{\text{LQ}} \not\sim p_{\text{HQ}}(\boldsymbol{x})$, the recovered latent may be pushed toward a less typical region of the prior space:

$$q\left(\hat{\boldsymbol{z}}_T^{\text{LQ}} \mid \boldsymbol{x}^{\text{LQ}}\right) \not\approx \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{3}$$

When an input image becomes less compatible with the learned HQ image distribution, this mismatch is reflected in the statistical atypicality of its inverted noise.

Based on this insight, we introduce the **Diffusion Inversion Gaussianity Score (DIGS)**, which quantifies the discrepancy between the distribution of an inverted latent and the Gaussian prior:

$$\mathrm{DIGS}(\boldsymbol{x}) = \mathcal{D}\left(q(\hat{\boldsymbol{z}}_T|\boldsymbol{x}), \mathcal{N}(\mathbf{0}, \boldsymbol{I})\right), \tag{4}$$

where $\mathcal{D}(\cdot,\cdot)$ measures the statistical deviation from the standard Gaussian prior. Accordingly, an HQ image is expected to produce a more typical inverted latent and hence a smaller DIGS, whereas an LQ image tends to induce a more atypical inverted latent.

Importantly, DIGS does not assume any specific degradation model. Instead of explicitly identifying whether an image suffers from blur, noise, compression, or other distortions, it asks a more fundamental question: *how naturally can this observation be explained by the Gaussian-to-HQ-image correspondence learned by the diffusion model?* This turns diffusion inversion from a reconstruction tool into a probe of image quality, enabling quality assessment directly through the statistics of the recovered generative noise.

3.2 DIGS: Marginal Gaussianity

Motivated by our hypothesis, we first characterize inversion atypicality through the standardized marginal distribution of latent coefficients. This captures deviations in distributional shape from the Gaussian prior while removing per-image location and scale effects.

Let $\boldsymbol{z} \in \mathbb{R}^{C\times H\times W}$ denote the terminal inversion latent, and let μ_z and σ_z denote its empirical mean and standard deviation. We define the absolute deviation from the Gaussian reference as

$$\Delta_k(\boldsymbol{z}) = \left| \frac{1}{n}\sum_{i=1}^{n} \left(\frac{z_i - \mu_z}{\sigma_z + \varepsilon}\right)^k - m_k^{\mathcal{N}} \right|, \tag{5}$$

where $m_k^{\mathcal{N}}$ denotes the k-th moment of $\mathcal{N}(0,1)$. We use $\varepsilon = 10^{-12}$ throughout for numerical stability.

The basic score uses the third and fourth orders. Skewness measures distributional asymmetry, whereas Pearson kurtosis measures the fourth central moment relative to the squared variance. Their Gaussian reference values are $m_3^{\mathcal{N}} = 0$ and $m_4^{\mathcal{N}} = 3$.

$$\mathrm{DIGS}(\boldsymbol{x}) = D_{\mathrm{marg}} = \Delta_3(\boldsymbol{z}) + \Delta_4(\boldsymbol{z}). \tag{6}$$

A lower DIGS indicates closer agreement with the standard-Gaussian marginal moments and is empirically associated with higher perceptual quality, refer to Fig. 4.

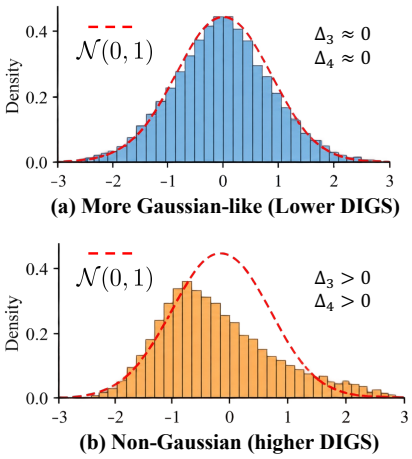


Figure 4: Marginal Gaussianity of inversion latents.

3.3 Higher-Order and Spatial Statistics

Only marginal gaussianity does not imply that the full coefficient distribution is Gaussian. Marginal statistics ignore channel identity and spatial organization in the inversion latent. We therefore introduce complementary higher-order and spatial statistics to capture these remaining structures, as shown in Figure 5.

Channel-wise higher-order deviation. To preserve this information, we characterize higher-order deviations of each latent channel independently and aggregate them across channels. Given the c-th latent channel $\boldsymbol{z}_c$, we measure the channel-wise higher-order deviation as

$$D_{ch} = \frac{1}{C}\sum\nolimits_{c=1}^{C} \Delta_k(\boldsymbol{z}_c). \tag{7}$$

In our implementation, we use a higher-order moment (we set $k = 5$) beyond skewness and kurtosis to capture channel-specific tail deviations. The absolute deviation avoids cancellation among channels with different asymmetry directions.

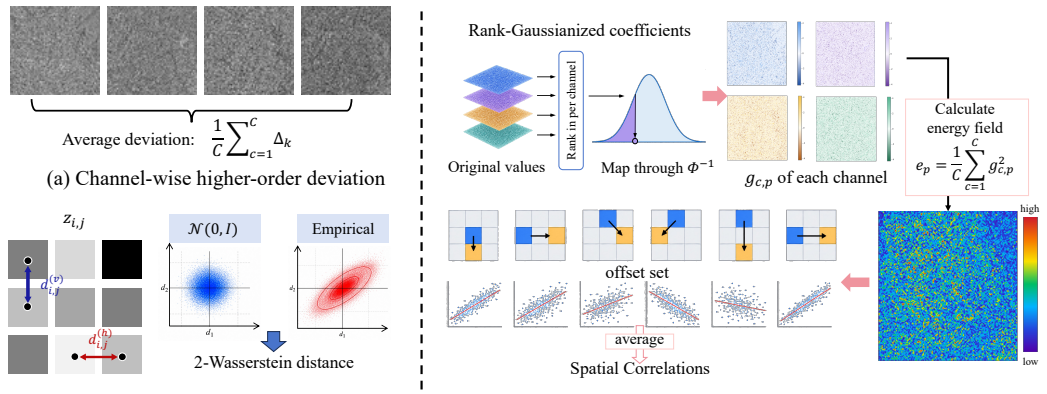


Figure 5: Computation pipeline of DIGS* from inversion latents, capturing marginal, channel-wise, and spatial quality-related statistics.

Whitened spatial differences. Since diffusion inversion preserves local spatial structures, we further measure whether neighboring latent differences remain consistent with the Gaussian difference prior. Specifically, we first whiten latent features $\tilde{\boldsymbol{z}}_p$ across channels within the selected region p and we have $\tilde{z}_p$. For each spatial direction $a \in \{h, v\}$, the normalized neighboring difference is $d_{i,j}^{(h)} = \frac{\tilde{z}_{i,j+1} - \tilde{z}_{i,j}}{\sqrt{2}}, d_{i,j}^{(v)} = \frac{\tilde{z}_{i+1,j} - \tilde{z}_{i,j}}{\sqrt{2}}$. Under independent standard Gaussian assumptions, $d_p^{(a)}$ also follows $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. We therefore estimate its empirical mean μ_a and covariance Σ_a, please refer to the details in the Supplementary. We compare the moment-matched Gaussian $\mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$ with $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ using the squared 2-Wasserstein distance and average over the two directions:

$$D_{\text{diff}} = \frac{1}{2} \sum_{a \in \{h,v\}} W_2^2(\mathcal{N}(\mu_a, \Sigma_a), \mathcal{N}(0, I)) = \frac{1}{2} \sum_{a \in \{h,v\}} \left[\|\mu_a\|_2^2 + \text{Tr}\left(\Sigma_a + I - 2\Sigma_a^{1/2}\right) \right]. \quad (8)$$

This score measures mean and covariance deviations of local difference vectors. Retaining the covariance matrix preserves cross-channel relationships among these differences that scalar flattening discards. The transport derivation is given in supplementary material.

Spatial energy dependence. While local difference statistics capture short-range variations of inversion latents, they do not explicitly characterize the global spatial organization of latent responses. We therefore measure the spatial dependence of latent energy to capture structural correlations preserved in the inversion process.

To reduce the influence of channel-wise magnitude variations and retain the relative spatial structure, we first Gaussianize within-image ranks separately for each channel, obtaining coefficients $g_{c,p}$ (see Supplementary) and the energy field $e_p = C^{-1} \sum_c g_{c,p}^2$. For six fixed spatial offsets $\mathcal{D}$, we compute

$$T_{\text{energy}} = \frac{1}{|\mathcal{D}|} \sum_{\delta \in \mathcal{D}} \text{Corr}_{p \in \mathcal{P}_\delta}(e_p, e_{p+\delta})^2, \quad (9)$$

where Corr is Pearson correlation. The squared correlation measures the strength of spatial dependence regardless of correlation direction, providing complementary information beyond marginal distribution and local difference statistics.

DIGS* combines complementary statistics from marginal, channel-wise, and spatial perspectives. D_{marg}, D_{ch}, and D_{diff} measure different types of deviations from the Gaussian prior, while T_{energy} captures the spatial organization preserved in inversion latents.

$$\text{DIGS}^*(\boldsymbol{x}) = \log \frac{\sqrt{T_{\text{energy}} + \varepsilon}}{(D_{\text{marg}} + \varepsilon)(D_{\text{diff}} + \varepsilon)(D_{\text{ch}} + \varepsilon)}. \quad (10)$$

The numerator reflects spatial dependence, whereas the denominator aggregates marginal, channel-wise, and local deviation terms. The logarithm transforms the multiplicative aggregation into an additive form and improves numerical stability. A higher $\text{DIGS}^*(\boldsymbol{x})$ indicates stronger observation–prior compatibility and is empirically associated with better perceptual quality.

Table 2: Comparison with existing no-reference IQA methods. The upper block reports supervised models trained on KonIQ-10k, and the lower block reports training-free and quality-label-free methods. Slashes indicate unreported in-domain results. Gray entries denote methods requiring quality-label training, and the best and second-performing entries are **bold** and underlined.

Quality-label training	Methods	KonIQ-10k		HRIQ		SPAQ	
		SROCC↑	PLCC↑	SROCC↑	PLCC↑	SROCC↑	PLCC↑
Yes	HyperIQA	Training set		0.482	0.798	0.844	0.841
Yes	TReS	Training set		0.745	0.736	0.418	0.417
Yes	MUSIQ	Training set		0.826	0.828	0.603	0.551
Yes	MANIQA	Training set		0.682	0.668	0.080	0.074
Yes	TOPIQ-NR	Training set		0.812	0.798	0.467	0.482
No	NIQE	0.527	0.474	0.610	0.593	0.504	0.472
No	IL-NIQE	0.494	0.476	0.270	0.280	0.692	0.637
No	PIQE	0.106	0.187	-0.022	0.142	0.064	0.072
No	CLIP-IQA	**0.685**	**0.685**	**0.725**	**0.709**	**0.726**	**0.720**
No	DIGS (ours)	0.516	0.482	0.612	<u>0.602</u>	0.671	0.664
No	DIGS* (ours)	<u>0.558</u>	<u>0.543</u>	<u>0.613</u>	0.601	<u>0.701</u>	<u>0.689</u>

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets and evaluation groups. We evaluate DIGS on three authentically distorted image collections: KonIQ-10k (10,073 images) (Hosu et al., 2020), HRIQ (1,120 images) (Huang et al., 2024), and SPAQ (11,125 smartphone photographs) (Fang et al., 2020). We use the supplied `MOS_zscore`, MOS, and overall MOS, respectively; all three targets increase with perceived quality. Each collection is evaluated independently on its annotated images.

Baselines. We compare DIGS with three groups of no-reference IQA methods. (i) Classical natural-scene-statistics approaches, including NIQE (Mittal et al., 2012b), IL-NIQE (Zhang et al., 2015), and PIQE (Venkatanath et al., 2015), which rely on handcrafted statistical priors. (ii) Frozen foundation-model based approaches, represented by CLIP-IQA (Wang et al., 2023), which directly extract quality-related signals from pretrained representations without IQA-specific optimization. (iii) Supervised deep IQA models, including HyperIQA (Su et al., 2020), TReS (Golestaneh et al., 2022), MUSIQ (Ke et al., 2021), MANIQA (Yang et al., 2022), and TOPIQ-NR (Chen et al., 2024). These models are trained with human quality annotations and are included only as an upper reference for learned IQA predictors rather than a directly comparable training-free setting.

Metrics and statistical protocol. We report Spearman rank-order correlation (SROCC) (Spearman, 1987) and Pearson linear correlation (PLCC) (Pearson, 1895). Raw DIGS is a *deviation*: lower values indicate better quality. Because human MOS increases with perceptual quality, we use $-\mathrm{DIGS}$ when computing correlation metrics. Importantly, the sign conversion is fixed before evaluation and does not involve any dataset-specific calibration or fitting. PLCC is computed directly, without a fitted nonlinear mapping. The scoring rule and direction are fixed across collections; neither MOS nor statistics from other test images enter any individual DIGS prediction.

Implementation details. Unless otherwise specified, we use the frozen SDXL base model[1] and perform 20-step deterministic DDIM inversion (Song et al., 2021a) with an empty text prompt. For each input image, the resulting terminal inversion latent is evaluated using DIGS or the corresponding statistical readout. We assess each readout by computing SROCC and PLCC between its per-image scores and the MOS provided by the dataset.

4.2 AGREEMENT WITH HUMAN QUALITY JUDGMENTS

We first investigate whether diffusion inversion statistics provide a reliable quality signal. As shown in Figure 1, DIGS exhibits a clear monotonic correlation with MOS across different authentic distortion datasets. The scatter distributions show that images with higher perceptual quality generally produce lower DIGS values, while lower-quality images tend to induce larger deviations from the Gaussian prior. This observation supports our hypothesis that the compatibility between an image observation and the pretrained diffusion prior can be reflected through the statistical properties of the recovered latent.

[1] `https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0`

Table 2 compares existing no-reference IQA methods. DIGS achieves the second-best performance among training-free approaches, consistently outperforming traditional natural-scene-statistics based methods. CLIP-IQA achieves slightly stronger correlation, likely benefiting from the rich semantic and visual priors encoded by large-scale image-language pretraining. In contrast, DIGS explores a different perspective by measuring the statistical compatibility between diffusion inversion latents and the intrinsic Gaussian prior, without relying on quality labels, reference images, or learned quality prompts. DIGS* further improves the performance by incorporating complementary channel-wise and spatial statistics, demonstrating the effectiveness of modeling inversion atypicality from multiple perspectives.

4.3 Sensitivity to Statistical Quality Measures

Ablation study. We conduct ablation studies to investigate the contribution of different statistical components in DIGS and DIGS*. As shown in Table 3, removing either skewness or kurtosis degrades the performance, while kurtosis contributes more significantly to the final quality correlation. This indicates that both asymmetric and tail distribution characteristics of inversion latents contain quality-related information, and their combination provides a more reliable measurement of latent Gaussianity. For DIGS*, we further evaluate the contribution of different complementary statistics. Individual spatial or channel-wise statistics provide weaker correlations when used alone, suggesting that no single statistic is sufficient to fully characterize inversion atypicality. However, progressively combining marginal, channel-wise, and spatial statistics consistently improves the performance.

Table 3: Component ablations. Entries are SROCC/PLCC; the full DIGS row uses the main fixed protocol.

Variant	KonIQ-10k		HRIQ	
	SROCC↑	PLCC↑	SROCC↑	PLCC↑
w/o kurtosis	0.299	0.292	0.445	0.429
w/o skewness	0.465	0.429	0.546	0.541
DIGS (ours)	0.516	0.482	0.612	0.602
D_{diff}	0.393	0.387	0.430	0.430
T_{energy}	0.283	0.302	0.470	0.444
D_{ch}	0.253	0.228	0.256	0.195
$D_{\text{marg}}+D_{\text{ch}}$	0.480	0.448	0.525	0.505
$D_{\text{diff}}+D_{\text{marg}}+D_{\text{ch}}$	0.509	0.490	0.564	0.550
DIGS* (ours)	0.558	0.543	0.613	0.601

Comparison of different measures. We compare DIGS with several commonly used statistical distances to examine whether generic distribution matching is sufficient for measuring inversion atypicality. In Figure 6, most distribution distances achieve weaker correlation with MOS compared with DIGS. While $\sqrt{JB}$ and D_{diff} provide meaningful quality signals by capturing certain aspects of non-Gaussianity or spatial deviation, they do not consistently characterize the quality-related variations in inversion latents. The superior performance of DIGS indicates that the shape statistics of standardized latent coefficients, especially skewness and kurtosis, provide a more suitable representation of diffusion-prior mismatch for no-reference image quality assessment.

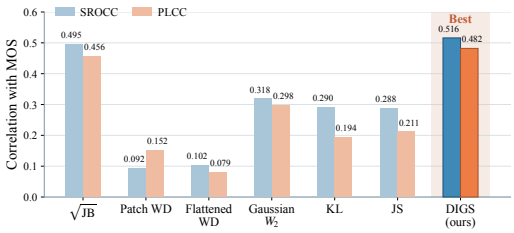


Figure 6: Comparison of alternative statistical quality measures on KonIQ-10k.

4.4 Sensitivity to Inversion Settings

Inversion operators and diffusion backbones. We further investigate the sensitivity of DIGS* to different inversion operators and diffusion backbones. As shown in Table 4, when using SDXL as the backbone, DDIM achieves the strongest correlation, while EDICT variants maintain comparable performance. In contrast, Euler inversion shows a larger performance degradation, suggesting that the choice of inversion trajectory affects the quality-related statistics extracted from terminal latents.

We further evaluate DIGS* with different backbones of diffusion models. Although the performance varies across models, several backbones, including PixelDiT, preserve meaningful correlation with human judgments. These results indicate that DIGS* is not limited to a single inversion configuration, while the quality of the underlying generative prior plays an important role in determining the strength of the measurement. Therefore, we retain the DDIM–SDXL combination as the default evaluation setting.

Inversion steps. The terminal latent depends on the discretization of the inverse trajectory. We therefore vary the number of DDIM inversion steps while keeping the image subset, SDXL back-

Table 4: Sensitivity of DIGS* to different inversion operators and diffusion backbones. The left panel compares inversion operators with SDXL fixed, while the right panel evaluates different diffusion backbones using the corresponding DIGS* readout.

(a) Inversion operators

Operator	SROCC	PLCC
DDIM (Song et al., 2021a)	0.558	0.543
Euler (Karras et al., 2022)	0.295	0.193
EDICT-x (Wallace et al., 2023)	0.541	0.524
EDICT-y (Wallace et al., 2023)	0.542	0.524
EDICT mean (Wallace et al., 2023)	0.542	0.524

(b) Diffusion backbones

Backbone	SROCC	PLCC
SDXL (Podell et al., 2024)	0.558	0.543
SD1.5 (Rombach et al., 2022b)	0.205	0.171
DiT-512 latent (Peebles & Xie, 2023)	0.308	0.285
ADM-256 (Dhariwal & Nichol, 2021)	0.326	0.296
PixelDiT (Yu et al., 2026)	0.534	0.518

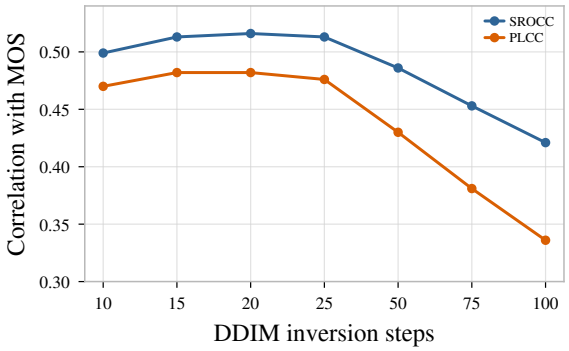


Figure 7: Sensitivity of inversion Gaussianity to the number of DDIM inversion steps on the KonIQ-10k subset. The quality correlation peaks at 20-th step.

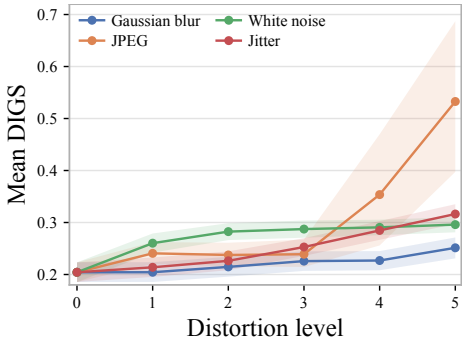


Figure 8: Effect of distortion severity. Curves show mean scores and shaded regions indicate the 95% confidence interval.

bone, empty prompt, and score readout fixed. In Figure 7, the correlation initially improves as the number of inversion steps increases from 10 to 20, while further increasing the steps leads to a gradual performance decrease. This indicates that the inversion process requires an appropriate trajectory length to preserve quality-related statistical characteristics while mapping the observation into the diffusion latent space. We therefore adopt 20-step DDIM inversion as the default setting.

4.5 Effect of Different Degradations

We further investigate whether DIGS responds consistently to different levels of image degradation. Using fixed-content images from KADID-10K (Lin et al., 2019), we progressively increase the severity of Gaussian blur, JPEG compression, white noise, and jitter, and measure the DIGS scores. From Figures 2&8, DIGS generally increases with distortion severity across different degradation types, indicating that stronger degradations lead to larger deviations from the diffusion prior. Among them, Gaussian blur, white noise, and jitter exhibit clear monotonic trends, while JPEG compression shows a similar overall increase with slight decreases at low severity levels and fluctuations at high severity levels. These results demonstrate that DIGS can capture progressive quality degradation without relying on predefined distortion categories.

5 Conclusion

In this work, we reveal a previously unexplored property of diffusion inversion: the Gaussianity of recovered noise provides a measurable signature of the compatibility between an image and a learned generative prior. Based on this finding, we introduce DIGS, a training-free and quality-label-free measure that quantifies image–prior compatibility through the statistical deviation of inversion noise from the intrinsic Gaussian prior, and further extend it with complementary latent statistics in DIGS*. Extensive experiments demonstrate that inversion Gaussianity consistently correlates with perceptual quality across diverse distortions and diffusion configurations. Our results establish diffusion inversion as a new measurement space for understanding how observations interact with generative priors, suggesting broader opportunities for probing the internal representations of foundation generative models.

AI USE STATEMENT

In this work, we have not used generative AI tools for generating synthetic data sets, proposing or refining hypotheses, developing theoretical models or conceptual frameworks, designing or providing feedback on research methodology or experiments, implementing methods, or assisting with translation. Formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, supporting qualitative and thematic data analysis, and cleaning and reformatting datasets do not apply to this work. Additionally, we used generative AI tools for polishing the writing. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Our experiments use only publicly available benchmark datasets (KonIQ-10k, HRIQ, SPAQ and KADID-10k) and publicly available super-resolution outputs; no new dataset is released. The preference experiment in the supplementary materials involved 30 adult volunteers ranking super-resolution outputs of the same inputs. Participation was voluntary; the task contained no sensitive content, and no personal or identifying information was collected. The study involved no vulnerable populations and posed minimal risk. We do not foresee direct harmful uses of the proposed score.

REPRODUCIBILITY STATEMENT

We support reproducibility by providing the complete mathematical definitions of DIGS and DIGS* in Section 3. Section 4 describes the public datasets and quality annotations, evaluation metrics, correlation directions, pretrained diffusion models, and the default 20-step deterministic DDIM inversion protocol. The supplementary material provides implementation-level definitions and derivations for channel whitening, neighboring latent differences, Gaussian transport, rank Gaussianization, and spatial energy dependence, including the fixed spatial offsets. It also documents the mask threshold, image-resizing procedure, baseline implementations, dataset- and configuration-specific statistical readouts, inversion-operator and diffusion-backbone comparisons, and the runtime and memory measurement protocol. The protocol for the super-resolution preference study, including input preparation and aggregation of participant rankings, is also described in the supplementary material. DIGS requires no quality-label training or target-dataset fine-tuning, and no fitted nonlinear mapping is used when computing the reported correlations. Upon acceptance, we will publicly release the complete code repository to facilitate further reproduction and research.

REFERENCES

Žiga Babnik, Peter Peer, and Vitomir Štruc. ediffiqa: Towards efficient face image quality assessment based on denoising diffusion probabilistic models. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE Transactions on Image Processing*, 2024.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 2021.

Wenkai Dong, Song Xue, Xiaoyue Duan, and Shumin Han. Prompt tuning inversion for text-driven image editing using diffusion models. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. Perceptual quality assessment of smartphone photography. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

Honghao Fu, Yufei Wang, Wenhan Yang, Alex C Kot, and Bihan Wen. Dp-iqa: Utilizing diffusion prior for blind image quality assessment in the wild. *arXiv preprint arXiv:2405.19996*, 2024.

Daniel Garibi, Or Patashnik, Andrey Voynov, Hadar Averbuch-Elor, and Daniel Cohen-Or. Renoise: Real image inversion through iterative noising. In *European Conference on Computer Vision*, 2024.

S. Alireza Golestaneh, Saba Dadsetan, and Kris M Kitani. No-reference image quality assessment via transformers, relative ranking, and self-consistency. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 2020.

Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 2020.

Huang Huang, Qiang Wan, and Jari Korhonen. High resolution image quality database. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024.

Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2014.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 2022.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *IEEE/CVF International Conference on Computer Vision*, 2021.

Xudong Li, Yan Zhang, Yunhang Shen, Ke Li, Runze Hu, Xiawu Zheng, and Sicheng Zhao. Feature denoising diffusion model for blind image quality assessment. In *AAAI Conference on Artificial Intelligence*, 2025.

Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *International Conference on Quality of Multimedia Experience*, 2019.

Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 2012a.

Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 2012b.

Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 2010.

Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 2011.

Zhaoqing Pan, Feng Yuan, Jianjun Lei, Yuming Fang, Xiao Shao, and Sam Kwong. Vcrnet: Visual compensation restoration network for no-reference image quality assessment. *IEEE Transactions on Image Processing*, 2022.

Karl Pearson. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 1895.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *IEEE/CVF International Conference on Computer Vision*, 2023.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022a.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022b.

Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Transactions on Image Processing*, 2012.

Shreshth Saini, Ru-Ling Liao, Yan Ye, and Alan Bovik. Lgdm: Latent guidance in diffusion models for perceptual evaluations. In *International Conference on Machine Learning*, 2025.

Qingbing Sang, Qian Li, Lixiong Liu, Zhaohong Deng, Xiaojun Wu, and Alan C Bovik. No-reference image quality assessment leveraging genai images. *IEEE Transactions on Image Processing*, 2025.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b.

Charles Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 1987.

Łukasz Staniszewski, Łukasz Kuciński, and Kamil Deja. There and back again: On the relation between noise and image inversions in diffusion models. In *International Conference on Learning Representations*, 2026.

Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

Narasimhan Venkatanath, Debbabi Praneeth, S Channappayya Sumohana, S Medasani Swarup, et al. Blind image quality evaluation using perception based features. In *National Conference on Communications*, 2015.

Bram Wallace, Akash Gokul, and Nikhil Naik. Edict: Exact diffusion inversion via coupled transformations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI Conference on Artificial Intelligence*, 2023.

Zhaoyang Wang, Bo Hu, Mingyang Zhang, Jie Li, Leida Li, Maoguo Gong, and Xinbo Gao. Diffusion model-based visual compensation guidance and visual difference analysis for no-reference image quality assessment. *IEEE Transactions on Image Processing*, 2025.

Peng Xing, Ning Wang, Jianbo Ouyang, and Zechao Li. Inv-adapter: Id customization generation via image inversion and lightweight parameter adapter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Jingtao Xu, Peng Ye, Qiaohong Li, Haiqing Du, Yong Liu, and David Doermann. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 2016.

Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Transactions on Image Processing*, 2014.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022.

Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2012.

Yongsheng Yu, Wei Xiong, Weili Nie, Yichen Sheng, Shiqiu Liu, and Jiebo Luo. Pixeldit: Pixel diffusion transformers for image generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Yan Zeng, Masanori Suganuma, and Takayuki Okatani. Inverting the generation process of denoising diffusion implicit models: Empirical evaluation and a novel method. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2025.

Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 2015.

Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.