Graph-Represented Distribution Similarity Index for Full-Reference Image Quality Assessment

Wenhao Shen, Mingliang Zhou, Jun Luo, Zhengguo Li *Fellow, IEEE*, Sam Kwong *Fellow, IEEE*

***Abstract*—In this paper, we propose a graph-represented image distribution similarity (GRIDS) index for full-reference (FR) image quality assessment (IQA), which can measure the perceptual distance between distorted and reference images by assessing the disparities between their distribution patterns under a graph-based representation. First, we transform the input image into a graph-based representation, which is proven to be a versatile and effective choice for capturing visual perception features. This is achieved through the automatic generation of a vision graph from the given image content, leading to holistic perceptual associations for irregular image regions. Second, to reflect the perceived image distribution, we decompose the undirected graph into cliques and then calculate the product of the potential functions for the cliques to obtain the joint probability distribution of the undirected graph. Finally, we compare the distances between the graph feature distributions of the distorted and reference images at different stages; thus, we combine the distortion distribution measurements derived from different graph model depths to determine the perceived quality of the distorted images. The empirical results obtained from an extensive array of experiments underscore the competitive nature of our proposed method, which achieves performance on par with that of the state-of-the-art methods, demonstrating its exceptional predictive accuracy and ability to maintain consistent and monotonic behaviour in image quality prediction tasks. The source code is publicly available at the following website[1].**

***Index Terms*—Image quality assessment, full reference, graph convolutional neural network, graph distribution**

I. Introduction

Image quality assessment (IQA) is a fundamental computer vision task that mimics the subjective image quality perception process of humans and is crucial for image compression, enhancement, and restoration [1]–[3]. In contrast to the no-reference method, which obtains quality results directly from the content of an image [4]–[6], full-reference IQA (FR-IQA) acquires quality assessment results by determining the difference between the distorted image and the reference image; this difference can be regarded as a measurement of the distance between these images. However, because of the complicated characteristics of the human visual system (HVS), this measurement strategy is still very challenging.

Objective FR-IQA approaches are usually executed by contrasting the pixels or patches in the distorted image with those in the reference image. The mean squared error (MSE) and peak signal-to-noise ratio (PSNR) have been considered standard full-reference indices, but due to their poor correlations with the subjective human perception process, the structural similarity index measure (SSIM) [7] and its related extensions [8]–[10], which better reflect perceptual quality, have become more widely accepted standards in practice. Feature engineering-based methods [11]–[13] can be divided into two steps: local feature extraction and perceptual global pooling techniques. Recently, numerous deep learning-based approaches have been proposed [14], [15] to achieve the goal of 'consistency with human perception' in terms of texture, structure, brightness, contrast, etc., which are essential components of the image content perception process implemented by the HVS [16]. As images are conventionally represented as pixel matrices arranged in a grid-like structure, many deep learning-based FR-IQA methods [16]–[18] rely on pixel block-level feature extractors, such as convolutional neural networks (CNNs) [19], [20]. Additionally, methods employing vision transformers (ViTs) [21]–[23] treat a two-dimensional image as a sequence of image blocks, enabling the establishment of rich visual relationships within the image content [24], [25].

Although many FR-IQA approaches have been recently proposed, several persistent issues warrant attention. First, a significant challenge lies in the fact that most methods tend to model specific distortion types. Thus, they lack the capacity to comprehensively characterize the HVS, consequently resulting in limited generalizability. In the context of FR-IQA, the responsiveness of a model to the distribution of image content is indicative of its ability to perceive both objects and distortion. Consequently, calculating the distance and dissimilarity between the distributions of a distorted image and its corresponding reference image serves as a means for assessing perceptual quality. Second, many of the existing methods often prioritize local perspectives rather than global perspectives. The spatial characteristics within images are

This work was supported in part by the National Natural Science Foundation of China under Grant 62176027; in part by the Key Projects of Basic Strengthening Plan under Grant 2022-JCJQ-ZD-018-11; in part by the Chongqing Talent under Grant cstc2024ycjh-bgzxm0082; in part by the Joint Equipment Pre Research and Key Fund Project of the Ministry of Education under Grant 8091B012207; in part by the Human Resources and Social Security Bureau Project of Chongqing under Grant cx2020073; in part by the Guangdong Oppo Mobile Telecommunications Corporation Ltd., under Grant H20221694; in part by the Hong Kong GRF-RGC General Research Fund under Grant 11209819 and Grant CityU 9042816; in part by the Hong Kong GRF-RGC General Research Fund under Grant 11203820 and Grant CityU 9042598; and in part by the Hong Kong Innovation and Technology Commission, InnoHK Project Centre for Intelligent Multidimensional Data Analysis (CIMDA).

Wenhao Shen and Mingliang Zhou are with the School of Computer Science, Chongqing University, Chongqing 400030, China (e-mail: shenwenhao163@163.com, mingliangzhou@cqu.edu.cn).

Jun Luo is with the School of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400030, China (e-mail: luojun@cqu.edu.cn).

Zhengguo Li is with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632, Singapore (e-mail: ezgli@i2r.a-star.edu.sg).

Sam Kwong is affiliated with the Computational Intelligence, Lingnan University, Hong Kong (email: samkwong@ln.edu.hk).

Corresponding author: Mingliang Zhou.

[1]https://github.com/Land5cape/GRIDS

inherently nonstationary, and the influence of superimposed noise information on the perceived visual quality is also nonstationary. Therefore, the aggregation of spatially local features may not consistently correspond to the overall perceptual quality of the distorted image. Third, the forms of objects in images are frequently highly irregular, making it challenging to effectively represent their correlations using traditional grid or sequence-like representations. These issues collectively underscore the complexity of FR-IQA problems and the need for novel methodologies to comprehensively address them.

To address these problems, we present an FR-IQA method that can indicate the perceptual quality distance between images by determining the difference between their graph distributions. Various factors, such as diverse lighting conditions, textures, and distortions, give rise to distinct image distributions. An image with a size of $H \times W$ obeys a distribution with certain $H \times W$-dimensional random variables. Since the dimensionality of these random variables is too high and they are not independent of each other, dimensionality reduction methods such as principal component analysis (PCA) have been proposed; these methods can approximate the original image with more important low-dimensional variables. Our approach exploits the strong spatial correlations among pixels by representing an image as a joint distribution of random variables corresponding to a selection of patches in the given image, where a graph probability model is utilized to establish the distribution relationships between these patches. Specifically, our method is constructed by utilizing a graph convolutional network (GCN) variant [26]–[28] that obtains distribution measurements over feature maps. Graph models offer advantages over CNN models because of their flexibility for use with irregular data and their ability to capture complex relationships, interpretability, transfer learning, scalability, and applicability across diverse domains. These methods excel when addressing nongrid data, intricate connections, and multimodal information. The inherent flexibility of CNN models allows them to accommodate and mitigate distortions of varying types and severities, making them versatile choices for image assessment tasks. Furthermore, to assess the cumulative influence of the nonstationary spatial features and noise information within an image on is overall visual perceptual quality, we employ clique construction on the image graph structure to derive the joint probability distribution of the undirected graph. The main contributions of our study are as follows.

- We propose a hierarchical deep architecture that obtains the feature distribution of the input image content at multiple scales, where the images are presented as graph representations that are more versatile and effective for visual perception purposes. This vision graph is automatically generated from the image content and can establish holistic perceptual connections for irregular regions.
- To reflect the perceived image distribution, we decompose the undirected graph into cliques and then calculate the product of the potential functions for the cliques to obtain the joint probability distribution of the undirected graph.
- We compare the regional distances between the graph distributions of the distorted and reference images at different stages; thus, the distortion distribution measurements derived from different graph model depths are combined to obtain the perceived visual quality of the distorted images.

This paper is structured as follows. In section II, we discuss the existing work related to objective FR-IQA methods and GCNs. In section III, we present a detailed description of the proposed FR-IQA method. The utilized experimental setups and the results acquired on public image quality databases are described in section IV. Finally, the conclusions are presented in section V.

II. Related Work

A. FR-IQA Methods

Feature engineering-based FR-IQA studies focus on designing features that can accurately model the image quality perception process of the HVS, which relies on prior knowledge or assumptions regarding the HVS. Assuming that the HVS is highly adaptive for extracting structural information from visual scenes, Wang *et al.* [7] proposed the SSIM, which represents a landmark in the field of IQA. Subsequently, a series of improved algorithms were proposed [8]–[10]. The multiscale SSIM (MS-SSIM) [8] incorporates more flexibility into the variations exhibited by observation conditions to calibrate the parameters that are relatively important at different scales. Wang *et al.* [9] proposed a wavelet domain index that measures the magnitude and phase changes in local wavelet coefficients caused by image distortion. The gradient-based SSIM [10], on the other hand, was proven to be more compatible with the HVS than the original SSIM for predicting the quality of blurred image. Considering that image gradients can sensitively reveal distortion, Xue *et al.* [29] proposed a gradient magnitude similarity deviation method that utilizes the global changes exhibited by gradient-based local quality maps to predict overall image quality. Zhang *et al.* [30] introduced a feature similarity approach that utilizes phase congruency and gradient magnitude information as the low-level characteristics of the HVS image interpretation process. Wang *et al.* [12] combined information content weighting with the MS-SSIM by relying on the assumption that the optimal perceptual weight should be proportional to the local information content to obtain a competitive perceptual quality metric. Ni *et al.* [31] proposed an edge similarity index that extracts and utilizes edge features in terms of edge contrast, edge width, and edge orientation information; these features are combined using an edge width aggregation strategy. Sun *et al.* [32] introduced a superpixel-based image quality assessment approach that considers luminance, chromaticity, and gradient similarities. Tang *et al.* [33] proposed a method utilizing multifeature extraction in both the spatial and frequency domains; this approach combines gradient magnitude and phase coherence information to characterize the distortion of a local structure. Bae *et al.* [34] introduced a structural contrast quality index that utilizes structural contrast information to characterize the perceived local and overall quality levels of diverse features with structural artificiality. Ahar *et al.* [35] performed sparse

coding based on the Fourier transform, enabling image quality evaluations to be performed on the basis of a correlation metric ranked by the amplitudes of the sparse coefficients. Fu *et al.* [36] described the edge information of distorted screen content images and reference images at two different scales using multiscale Gaussian differencing.

B. Deep Learning-Based FR-IQA Methods

Recently, data-driven deep learning-based approaches that can spontaneously learn HVS behaviours from enormous amounts of data have started to emerge. Kim *et al.* [37] introduced a CNN-based approach that is capable of solving for visual weights with an understanding of the available information, and the behaviour of the HVS is obtained from the underlying data distribution. Gao *et al.* [38] proposed a deep similarity index that measures the local similarities between features, and the global perceived quality is determined by pooling local quality indices. Bosse *et al.* [39] introduced a deep architecture that jointly learns the local quality and local weights through an overall framework. Wu *et al.* [40] introduced a structure with two CNN modules to identify the region of interest and predict the quality of ultrasound images by evaluating the descriptions of their key structures. By dynamically generating perceptual areas, Kim *et al.* [16] introduced a method that is responsive to distortion; their approach includes two adaptable error representation streams and a quality ensemble that leverages visual acuity. Ma *et al.* [41] proposed an end-to-end IQA structure that exhibits intrinsic invariance to geometric distortions to directly predict image quality. To evaluate the quality of an image compromised by distortions generated from various image processing algorithms, Ahn *et al.* [42] proposed a method that encodes distortion types and thus predicts distortion sensitivity. Sim *et al.* [43] proposed the mean and deviation of deep and local similarity, where the mean and deviation pooling operations were adopted to aggregate the local quality of the similarity map obtained from CNN features. Zhang *et al.* [44] classified screen content images into distinct regions in a fuzzy manner, and the predicted quality was determined by a quality degradation model with an edge structure, quality fusion, and region size adaptation.

C. Graph Neural Networks

GCNs, which are initially proposed by Bruna *et al.* [45], are common convolutional techniques in deep learning that are applied to data with graph structures. GCNs demonstrate strong expression capabilities in terms of learning graph representations, achieving excellent performance across various tasks and applications. Currently, GCNs are predominantly used to address data problems in non-Euclidean space and explore the potential relationships among data. With the continuous development of GCNs, they have been applied in diverse fields, such as recommendation systems [46], [47], natural language processing [48], [49], and computer vision [50], [51]. GCNs are primarily applied in computer vision for processing point cloud (PC) data because they excel at handling inherent graphical structure information [51], [52]. Recently, efforts have been dedicated to employing GCNs for solving tasks such as image classification [53], object detection [50], and semantic segmentation [52]. However, the challenge of employing a GCN for image quality assessment lies in the effective utilization of the non-Euclidean structures that are present in Euclidean space data, particularly in cases involving two-dimensional images, to derive efficient quality representations.

Distinct from the previously developed deep learning-based IQA methods, the proposed graph-represented image distribution similarity (GRIDS) method is in line with the available statistics-based models for visual images and the efficient coding hypothesis for neurons [54]. From the above two perspectives, a distortion in an image interferes with the natural distribution pattern of the original image, leading to perceptual variations in the distortions imposed on different contents. Our method predicts the fundamental phenomenon regarding the strong spatial correlations among pixels, and by modelling the perceptual distribution of an image as a graph probability model, the implicit relationships between image contents can be revealed while establishing an overall joint perceptual connection, which is intrinsically aligned with the perception of image distortion by humans.

III. Method

A. Framework

The objective of our approach is to map the local characteristics of the reference and distorted images, namely, I_{ref} and I_{dis}, respectively, to a perceptual global construct. The framework of our approach is presented in Figure 1, and it comprises three main operations: graph construction, graph encoding, and graph distribution measurement. The graph distribution measurement process consists of two key components: the construction of cliques for graph distribution representation and the subsequent perceptual distance measurement step. Notably, the network parameters that govern the processing of both the reference and distorted images are shared, ensuring a consistent and unified processing approach across both images. Before beginning the initial stage of our graph distribution-based GCN encoder, the input images are processed through a series of CNN layers. This approach not only enhances the feature dimensionality but also downsamples the input dimensions. Furthermore, the downsampling operation is implemented at the concluding block of each stage for effectively extracting features and performing dimensionality reduction. This process facilitates a comprehensive analysis of the image features that are present at various scales and regions, contributing to a nuanced understanding of the perceptual quality of the given images.

Image graph construction. The aim of graph construction is to establish perceptual relationships between the contents of different areas in an image. Since certain correlations exist among the contents of an image, the reference image I_{ref} and distorted image I_{dis} with widths and heights of H and W, respectively, are cropped into several patches via a series of convolution operations. For each image, we construct an

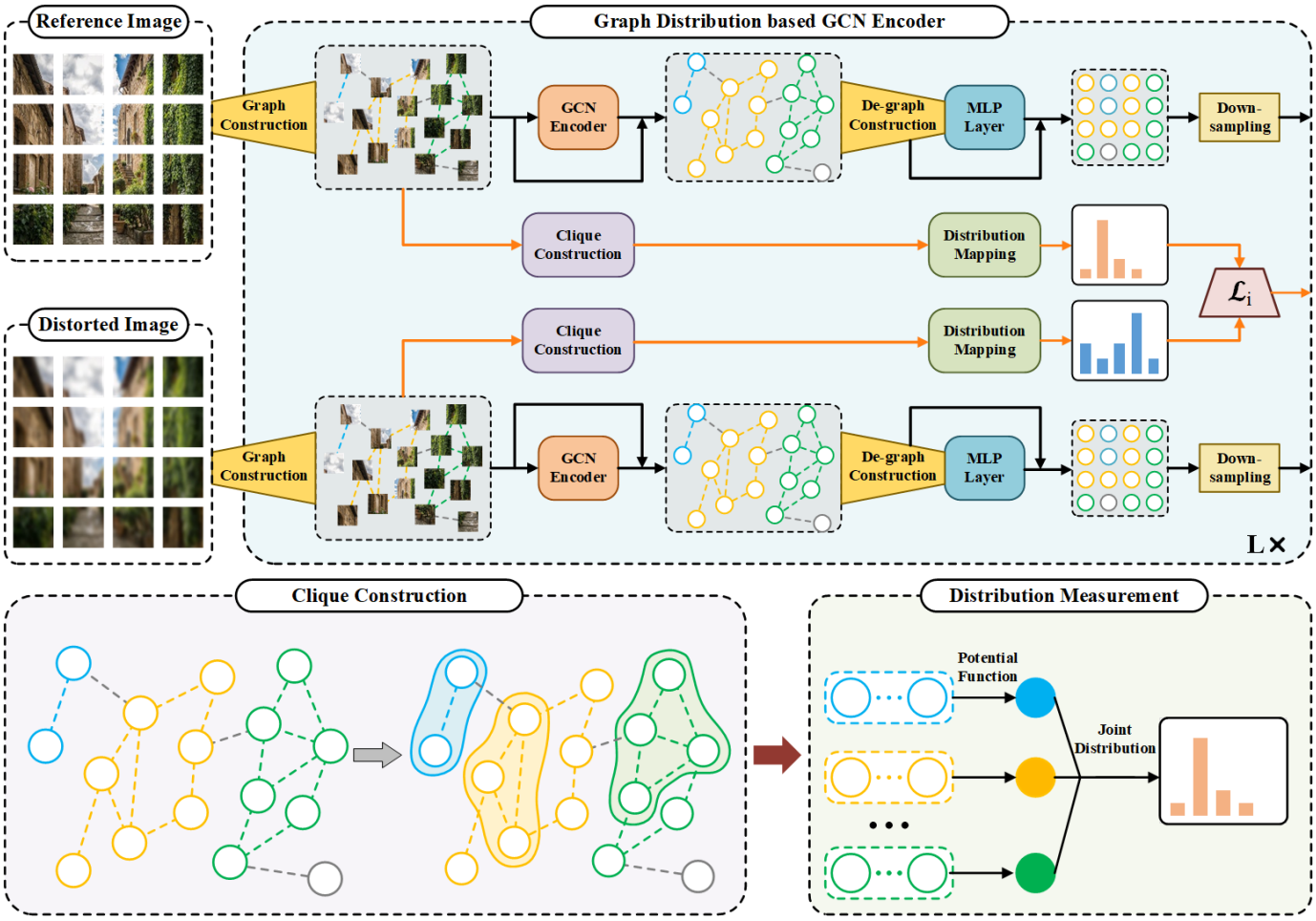


Fig. 1. Illustration of the proposed method. The input reference and distorted images undergo multiple network stages, each with two branches: one branch involves graph construction, graph convolution encoding, graph deconstruction, skip connections, and MLP layers; the other branch includes the graph clique construction process and the graph distribution measurement procedure.

undirected graph model G to describe the connections between them, and this model can be expressed as

$$G = \{V, E\}, \tag{1}$$

where V represents the set of vertices and E represents the set of edges. The relevance of these patches, or vertices, is described by the edge weights of the graph. More specifically, we build an adjacency matrix $\mathbb{A}$ for all vertices of an image, where the value of a_{ij} in $\mathbb{A}$ denotes the degree of correlation between vertices i and j. To preserve the position information of these vertices when constructing the graph, relative position encoding is incorporated into the adjacency matrix.

The graphical representation of an image, specifically its adjacency matrix, is not static but rather dynamically changes. Before each graph convolution operation, the graph representation of the image is reconstructed to capture finer-grained information. After each graph convolution operation, the graphical representation of the image is subsequently transformed back into a grid matrix representation. This transformation is succeeded via the application of a multilayer perceptron (MLP) for extracting additional features.

Our approach involves gradually reducing the initial high-resolution image in a stagewise manner, collectively constructing a hierarchical structure, as depicted in Figure 2. To elaborate further, each stage is composed of a series of blocks incorporating graph convolutions and MLPs. In the final block of each stage, a downsampling operation is used to connect spatially adjacent 2×2 patches, effectively expanding the receptive field of the network. To decrease the spatial dimensions of the given image while retaining its crucial visual details, we employ convolution kernels with sizes of 3×3 and strides of 2×2 during the downsampling process. This procedure enables the creation of new adjacency matrices with vertices representing the image patches at different scales in each stage.

Graph encoding. In each stage, a series of GCN encoders are employed to extract the features of the constructed graph. Akin to CNNs, GCNs extract rich features from each vertex and its neighbours; however, they differ in that the convolution in GCNs is implemented by an aggregation operation:

$$f_i^{(l+1)} = \phi^{(l)}(f_i^{(l)}, f_j^{(l)}; \mathbb{W}_i, \mathbb{W}_j \mid j \in \mathcal{N}^{(l)}(i)), \tag{2}$$

where $f_i^{(l)}$ and $f_j^{(l)}$ represent the features of vertices i and j at the l^{th} layer, respectively, $\mathbb{W}_i$ and $\mathbb{W}_j$ denote the weight matrices, $\mathcal{N}^{(l)}(i)$ indicates the neighbouring vertices of vertex

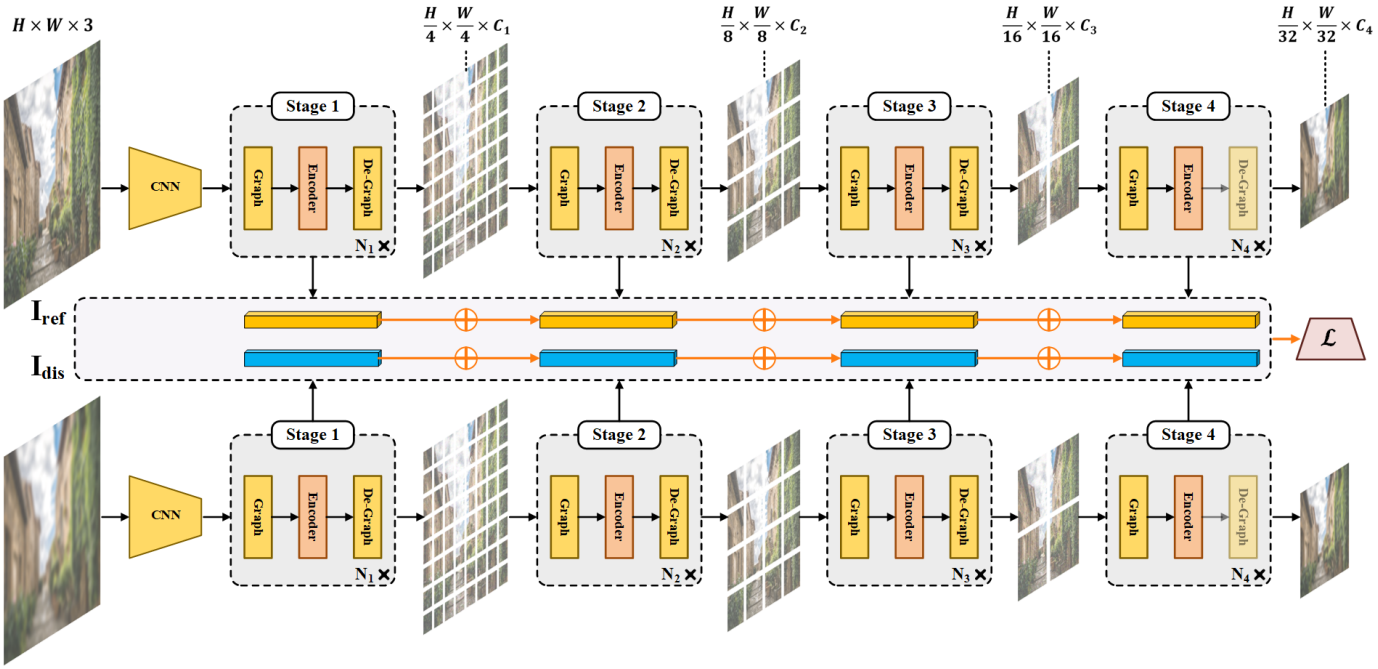


Fig. 2. Illustration of the hierarchical distance measurement strategy. Our approach reduces the initial high-resolution image in a stepwise manner, resulting in a hierarchical structure. The input reference and distorted images in each network stage acquire a feature vector representing the graph distribution. During each stage, a distance measurement method is employed to quantify the distribution differences between the reference and distorted images, culminating in a global graph-based similarity index.

i, and $\phi^{(l)}(\cdot)$ indicates the aggregation function. Specifically, we employ the max-relative GraphConv [27] method for aggregating neighbourhood relative features, which can be formulated as follows:

$$f_i^{(l+1)} = max(f_i^{(l)} - f_j^{(l)} \mid j \in \mathcal{N}^{(l)}(i)). \tag{3}$$

Subsequently, the features output by this layer are concatenated with $f_i^{(l)}$, after which they are passed through the MLP layers to obtain the corresponding aggregated features. These features are combined with skip connections. After the aggregation process, the updated feature of each vertex reflects the superimposed value of that vertex and its surroundings.

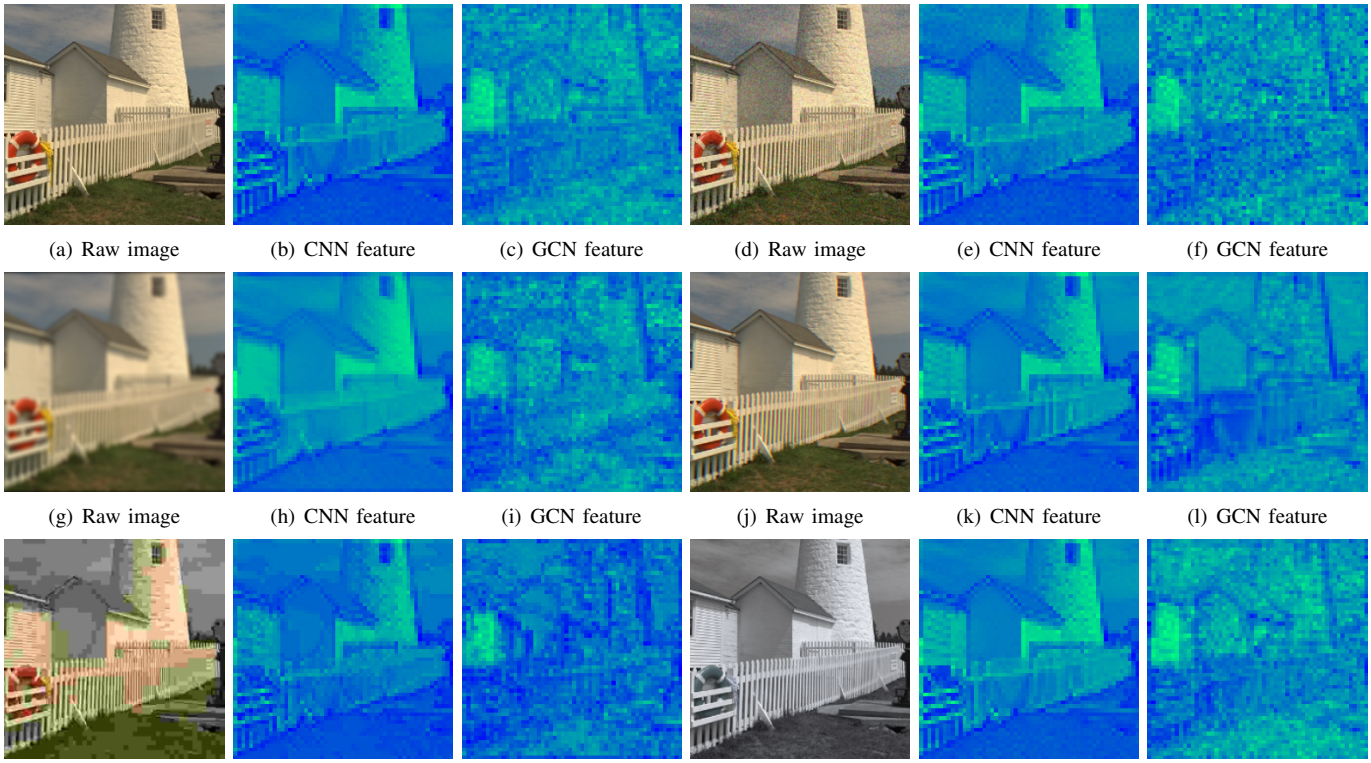
The graph G updates after all vertices are aggregated once, which can be formulated as

$$G^{(l+1)} = \mathcal{F}(G^{(l)}, \mathbb{W}^{(l)}), \tag{4}$$

where $G^{(l)}$ and $G^{(l+1)}$ denote the graphical representations at the l^{th} and $l+1^{th}$ layers, respectively, $\mathbb{W}^{(l)}$ reflects the weight matrix of $G^{(l)}$, and $\mathcal{F}(\cdot)$ denotes the progress of the update strategy.

The graph encoding network provides rich image feature representations, as illustrated in Figure 3. Compared to the feature maps of the original image, the feature maps of the distorted images effectively reflect the distortion information of the image. For example, in a case with an image possessing additive Gaussian noise, its feature map exhibits more high-frequency noise points, whereas for an image with Gaussian blur, the feature map shows a higher prevalence of low-frequency distributions. These maps underscore the efficacy of our GNN with respect to capturing distortion features.

Graph distribution measurement To determine the perceptual distance between the reference image and the distorted image, a graph distribution measurement scheme is implemented to evaluate the distribution distance between the multiscale graph structures of images. Our approach can be divided into two steps: graph distribution representation and perceptual distance measurement. After executing feature extraction using GCNs, we transform the graph feature maps into graph distributions. This conversion enables a more detailed understanding of the feature distribution across various scales and regions within the image. This fine-grained information is of utmost importance for image quality assessment tasks, as it aids in capturing subtle perceptual characteristics, ultimately resulting in more precise quality predictions. Essentially, this process can be viewed as mapping the original features to a higher-level abstract representation, thereby enhancing the perceptual quality estimation step. Moreover, the establishment of cliques is pivotal because it enables the segmentation of an undirected graph into locally connected components, resulting in a concise and interpretable representation of the image content distribution, which is a critical factor for precisely assessing quality. The visual representation in Figure 4 illustrates the connectivity of a graph under various scenes, confirming that the obtained graph structure and clusters can effectively aggregate similar patches in the corresponding image. This approach establishes a comprehensive perceptual relationship for the assessing the quality of different regions. Subsequently, we quantitatively inscribe the correlations between the variables in the parameter set (all patches of an image) and construct this set as a potential function $E(x)$. The joint probability distribution of the undirected graph can be expressed as the



(a) Raw image (b) CNN feature (c) GCN feature (d) Raw image (e) CNN feature (f) GCN feature

(g) Raw image (h) CNN feature (i) GCN feature (j) Raw image (k) CNN feature (l) GCN feature

(m) Raw image (n) CNN feature (o) GCN feature (p) Raw image (q) CNN feature (r) GCN feature

Fig. 3. Feature maps of images with different distortions. The raw images serve as the inputs for our approach, while the CNN feature maps represent the outputs obtained from the initial stage of our network, where dimensionality reduction is performed using a CNN. The GCN feature maps are the outputs of stage 1 after the graph construction and graph encoding processes are completed in our method. These feature maps are rescaled, and graph constructions are generated in the corresponding original positions. **(a)-(c):** Reference images. **(d)-(f):** Images with additive Gaussian noise distortion. **(g)-(i):** Images with Gaussian blur distortion. **(j)-(l):** Images with chromatic aberration distortion. **(m)-(o):** Images with JPEG compression distortion. **(p)-(r):** Images with changes in colour saturation.

product of the potential functions for the cliques decomposed from the graph. Hence, we compare the distribution differences between the reference and distorted images, and since such discrepancies are present in all phases of the model, we use a uniform metric to measure the distance between them.

B. Graph Distribution Representation

In the context of a graph G, we introduce the random variables associated with the vertices in V as $X = x_1, x_2, \cdots, x_B$. Among them, each x_i corresponds to a specific block, signifying the label attributed to that block, and B denotes the total number of blocks within each image. Given that adjacent pixels exhibit noncausal interactions, we construct correlations through an undirected graph representation-based probability model. The inclusion of a graph distribution representation in our methodology holds significant value, as it enables the capture of intricate relationships and the establishment of holistic perceptual connections, particularly for the irregular regions within an image. This capability further facilitates the integration of both semantic- and distortion-related information. To elucidate this process, we begin by decomposing the undirected graph into cliques. Subsequently, we calculate the product of the potential functions associated with these cliques. This step results in the formulation of a joint probabilistic graphical model for the undirected graph, as follows:

$$P(X) = \frac{1}{Z} \prod_{Q \in S} e^{E(x_Q)}, \tag{5}$$

where e is the natural constant, S denotes the set of all cliques, Q is one of the cliques in S, and Z denotes the partition function that normalizes the characters:

$$Z = \sum_{x} \prod_{Q \in S} e^{E(x_Q)}, \tag{6}$$

and $E(\cdot)$ represents the potential function:

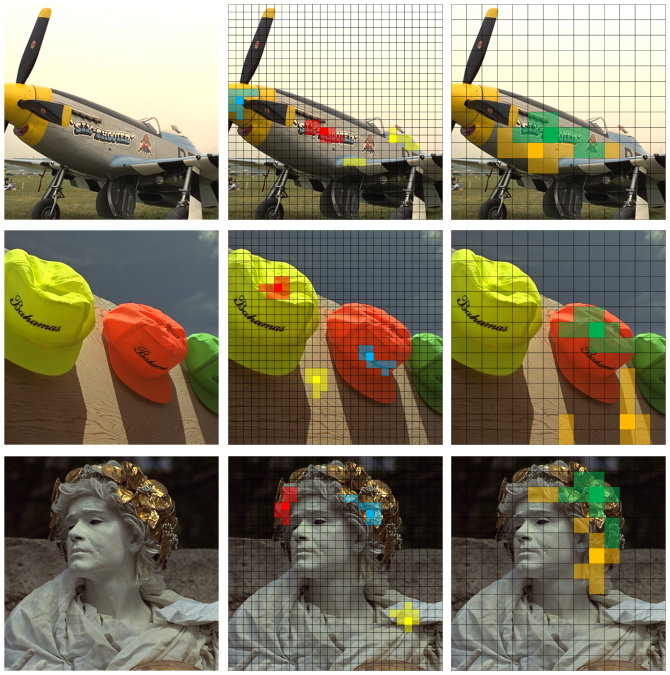
$$E(x_Q) = \sum_{i \in Q} \psi_1(x_i) + \sum_{i,j \in Q} \psi_2(x_i, x_j), \tag{7}$$

where the univariate part $\psi_1(x_i)$ measures the cost of the block i that takes the label x_i, and the bivariate component $\psi_2(x_i, x_j)$ measures the cost of simultaneously assigning the labels x_i and x_j to the pixel blocks i and j, respectively.

The function $\psi_2(x_i, x_j)$ can be expressed as the product of two factors:

$$\psi_2(x_i, x_j) = \rho(x_i, x_j) \cdot \xi(f_i, f_j), \tag{8}$$

where f_i represents the feature vector of block i. The function $\rho(\cdot)$ captures the relationships between different distribution pairs that assign penalties accordingly when different tags are assigned to blocks i and j with similar properties, and $\xi(f_i, f_j)$



(a) (b) (c)

Fig. 4. Graph connection visualization. **(a)** Original image. **(b)** Graph connections of the second stage. **(c)** Graph connections of the third stage. To yield a more explicit representation, we distinguish the connections of partial nodes with different colours, where the opaque node is the centre point.

computes the variability among the original features. Inspired by the formulation of the SSIM, we measure the variability in terms of the mean, standard deviation and covariance of each corresponding pair:

$$\xi(f_i, f_j) = 1 - \frac{(2\mu_i\mu_j + \tau_1)(2\delta_{ij} + \tau_2)}{(\mu_i^2\mu_j^2 + \tau_1)(\delta_i^2\delta_j^2 + \tau_2)}, \tag{9}$$

where $\mu_i, \mu_j, \delta_i, \delta_j$ and δ_{ij} represent the global means and standard deviations of blocks i and j, respectively, as well as their covariance. Additionally, τ_1 and τ_2 are constants utilized to prevent the denominator from being 0. The predicted distribution can be formulated as follows:

$$\begin{aligned} P(X) &= \frac{1}{Z}\prod_{Q\in S} e^{E(x_Q)} \\ &= \frac{1}{Z} e^{\sum_{Q\in S} E(x_Q)} \\ &= \frac{1}{Z} e^{\sum_{Q\in S}(\sum_{i\in Q}\psi_1(x_i) + \sum_{i,j\in Q}\psi_2(x_i,x_j))}. \end{aligned} \tag{10}$$

C. Perceptual Distance Measurement

As shown in Figure 2, the proposed perceptual distance measurement scheme is utilized in all stages of our network, from shallow to deep. We use the hierarchical network to compute the graph encoding as well as the distribution, where the distance measurement is utilized to measure the distribution variations between the reference image I_{ref} and the distorted image I_{dis}:

$$d(I_{ref}, I_{dis}) = \frac{1}{\sqrt{2}} ||\sqrt{p_{ref}} - \sqrt{p_{dis}}||_2, \tag{11}$$

Algorithm 1: Graph-represented image distribution similarity index

Input: the reference image I_{ref} and distorted image I_{dis} with sizes of $H \times W$
Output: the predicted quality score

Let M represent the number of stages and N_m denote the number of encoder blocks at stage m
for $m = 1$ ***to*** M **do**
  for $n = 1$ ***to*** N_m **do**
    Transform images I_{ref}^n and I_{dis}^n into graphs G_{ref}^n and G_{dis}^n
    Perform graph feature encoding via Eq. (4)
    Decompose the undirected graph into cliques
    Calculate the graph distribution via Eq. (5)
    Measure the feature distribution distance via Eq. (11)
    Deconstruct graphs G_{ref}^n and G_{dis}^n
    Process through the MLP layers
  end
  Downsample through convolution operations
end
Obtain the predicted image quality level via Eq. (12)
return quality score

where p_{ref} and p_{dis} denote the graph distribution representations of I_{ref} and I_{dis}, respectively. Thus, our proposed method combines the distribution measurements acquired from M different model stages:

$$D(I_{ref}, I_{dis}) = 1 - \frac{1}{M}\sum_{i=1}^{M} d_i(p_{ref}, p_{dis}). \tag{12}$$

In practice, we minimize the error between the predictions and the ground truths for quality assessment purposes; thus, our loss function is defined as follows:

$$\mathcal{L} = |D(I_{ref}, I_{dis}), Mos(I_{dis})|, \tag{13}$$

where $Mos(\cdot)$ indicates the ground truth of the mean opinion score (MOS).

The algorithmic procedure of our method can be expressed as Algorithm 1. To clarify the entire algorithmic process, we organize the formula symbols in the text and present them in Table I for reference. During the process, first, the images I_{ref} and I_{dis} with sizes of $H \times W$ are separately processed into patches with the same scale, and the patches of each image compose graphs according to their spatial locations and content similarity; these graphs are denoted as G_{ref} and G_{dis}, respectively. Next, we employ graph encoders to extract the image features represented by each graph. The graph structure features are then reconstructed into a two-dimensional matrix, and the feature representations are enriched by residual concatenation and MLP operations with downsampling in each stage to reduce the required computational effort. Simultaneously, the graph features encountered during this process are utilized for computing the distribution, as they encapsulate more information about the image content. Specifically, we decompose each graph into cliques, where the distribution

TABLE I
SUMMARY OF THE UTILIZED NOTATIONS.

Symbol	Description	Symbol	Description
I_{ref} and I_{dis}	Reference image and distorted image	H, W and C	Image height, width and number of channels
V and E	Vertices and edges	G	Graph consisting of V and E
$\mathbb{A}$	Adjacency matrix	i and j	Vertex index
a_{ij}	Adjacency value of vertices i and j in $\mathbb{A}$	l	Layer index
$f_i^{(l)}$	Features of vertex i in the l^{th} layer	$\mathbb{W}$	Weight matrix
$\mathcal{N}^{(l)}(i)$	Neighbour vertices of vertex i in the l^{th} layer	$\mathcal{F}(\cdot)$	Update progress of G
$\phi^{(l)}(\cdot)$	Aggregation function	$E(\cdot)$	Potential function
X	Random variable	x_i	Label attributed to block i
B	Number of image blocks	S	Set of all cliques
Q	Clique in S	Z	Partition function
M	Number of stages	N_m	Number of encoder blocks in stage m
Mos	Mean opinion score	τ_1 and τ_2	Constants
p_{ref} and p_{dis}	Graph distribution representations of I_{ref} and I_{dis}	K	Number of neighbours

of the entire graph can be expressed as the product of the normalized potential functions for the cliques. Finally, we compare the distances between the graph distributions of the reference and distorted images at different stages; thus, the distortion distribution measurements derived from different graph model depths are integrated to derive the perceived visual quality of the images. This comprehensive approach leverages graph-based representation, feature extraction, distribution measurement, and distance comparison to effectively assess image quality.

Notably, although some methods [55], [56] incorporate graph algorithms into quality assessment tasks, they mainly focus on assessing the quality of PC images rather than 2-dimensional images. In contrast, inspired by sequence-based techniques such as transformers (e.g., the ViT [24]), we adopt a different approach. We divide each image into patches with equal scales. The introduced graph structures serve the purpose of establishing connections between the contents of various regions within an image. This approach provides a unique image representation framework tailored to IQA challenges.

IV. Experimental Results

A. Experimental Settings

Implementation details. The complete network is optimized in an end-to-end manner by minimizing the discrepancy between the predicted values and the MOS. Our optimization strategy employs the weighted adaptive moment estimation (AdamW) [71] optimizer with a learning rate set to 0.0001, a weight decay rate of 0.05, and a path dropping rate of 0.3. Furthermore, we incorporate a warmup initialization phase followed by a cosine annealing-based learning rate schedule. In the training process, we maintain a batch size of 64 and continue training for a total of 200 epochs. To ensure fair comparisons, all experiments involving the compared methods are conducted on the same computational platform [2]. The design of our model employs a specific number of blocks in each stage, with the block count set as [2, 2, 6, 2]; these blocks are accompanied by channel sizes of [80, 160, 400, 640]. The number of neighbours, denoted as K, is set to 9. During the experimental phase, we normalize the MOS values for each database to align with their respective label ranges. This standardization process aids in achieving improved performance and facilitates equitable comparisons among different datasets.

Compared methods. In our study, we compare several representative and state-of-the-art methods, which encompass both feature engineering-based methods and deep learning-based methods, including the PSNR, SSIM [7], MS-SSIM [8], complex wavelet-SSIM (CW-SSIM) [72], visual saliency-based index (VSI) [64], most apparent distortion (MAD) [58], visual information fidelity (VIF) [63] criterion, feature similarity (FSIM) [30], gradient magnitude similarity deviation (GMSD) [29], structural contrast quality index (SCQI) [34], normalized Laplacian pyramid (NLPD) [65], DeepFL-IQA [66], DeepQA [67], weighted-average deep image quality measure (WaDIQaM) [39], perceptual image error assessment through pairwise preference (PieAPP) [14], just-noticeable-difference saliency channel attention residual (JND-SalCAR) model [68], learned perceptual image patch similarity (LPIPS) [73], deep image structure and texture similarity [17], conditional knowledge distillation network [74], image quality transformer (IQT) [21], attention-based hybrid image quality assessment network (AHIQ) [69], unifying dual-attention and Siamese transformer network (UniDASTN) [18] and joint semisupervised and positive unlabelled learning method (JSPL) [23]. To assess the effectiveness of these methods, we rely on the Pearson linear correlation coefficient (PLCC) and the Spearman rank correlation coefficient (SRCC) as the primary performance metrics. The PLCC quantifies prediction accuracy, while the SRCC measures monotonicity by assessing the rank correlation between the predicted scores and the labels derived from the corresponding MOSs.

Database. Six benchmark IQA databases are utilized to verify the validity of our method: the laboratory for image &

[2]PyTorch is used on a system with an Intel Core i9-10900 central processing unit (CPU), a 24-GB NVIDIA RTX3090 graphics processing unit (GPU) and 64 GB of random access memory (RAM).

TABLE II
SUMMARY OF THE SIX UTILIZED IMAGE QUALITY DATABASES.

Dataset	LIVE [57]	CSIQ [58]	TID2008 [59]	TID2013 [60]	KADID-10k [61]	PIPAL [62]
Reference images	29	30	25	25	81	250
Distorted images	779	866	1,700	3,000	10,125	25,850
Distortion types	5	6	17	24	25	40
Image resolution	-	512 × 512	512 × 384	512 × 384	512 × 384	288 × 288
Score range	[0, 100]	[0, 1]	[0, 9]	[0, 9]	[1, 5]	[868, 1857]

TABLE III
RESULTS OF A PERFORMANCE COMPARISON CONDUCTED ON FIVE BENCHMARK DATASETS.

Method	LIVE [57]		CSIQ [58]		TID2008 [59]		TID2013 [60]		KADID-10k [61]	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
PSNR	0.791	0.807	0.781	0.801	0.507	0.524	0.663	0.686	0.667	0.672
SSIM [7]	0.847	0.851	0.819	0.832	0.625	0.624	0.665	0.627	0.780	0.778
MS-SSIM [8]	0.886	0.903	0.864	0.879	0.841	0.853	0.785	0.729	0.835	0.834
VIF [63]	0.948	0.952	0.898	0.899	0.798	0.749	0.771	0.677	0.676	0.669
MAD [58]	0.904	0.907	0.934	0.932	0.814	0.833	0.803	0.773	0.829	0.827
FSIM [30]	0.910	0.920	0.902	0.915	0.874	0.883	0.876	0.851	0.850	0.850
VSI [64]	0.877	0.899	0.912	0.928	0.871	0.895	0.898	0.894	0.875	0.876
GMSD [29]	0.909	0.910	0.938	0.939	0.878	0.890	0.858	0.804	0.847	0.846
NLPD [65]	0.882	0.889	0.913	0.925	0.865	0.877	0.832	0.799	0.819	0.820
SCQI [34]	0.937	0.948	0.927	0.943	0.895	0.908	0.907	0.905	0.853	0.854
DeepFL-IQA [66]	0.978	0.972	0.946	0.930	0.875	0.867	0.876	0.858	0.938	0.936
DeepQA [67]	0.982	0.981	0.965	0.961	0.951	0.947	0.947	0.939	0.910	0.912
WaDIQaM-FR [39]	0.980	0.970	0.967	0.962	0.953	0.945	0.946	0.940	0.935	0.931
PieAPP [14]	0.986	0.977	0.975	0.973	0.956	0.951	0.946	0.945	0.832	0.830
JND-SalCAR [68]	0.987	0.984	0.977	0.976	0.957	0.956	0.956	0.949	0.960	0.959
AHIQ [69]	0.989	0.984	0.978	0.975	0.964	0.957	0.968	0.962	0.958	0.961
VTAMIQ [70]	0.984	0.982	0.982	0.979	0.951	0.948	0.954	0.950	0.964	0.963
JSPL [23]	0.980	0.983	0.977	0.970	0.948	0.950	0.940	0.949	0.961	0.963
GRIDS (ours)	0.988	0.986	0.979	0.976	0.961	0.954	0.959	0.953	0.967	0.965

Performance comparison between the results obtained by our method and the state-of-the-art methods on five benchmark datasets. In each column, the best, second-best and third-best results are highlighted in red, green and blue, respectively.

video engineering (LIVE) image quality assessment database [57], the categorical subjective image quality (CSIQ) database [58], the 2008 Tampere image database (TID2008) [59], the 2013 Tampere image database (TID2013) [60], the Konstanz artificially distorted image quality database (KADID-10k) [61] and the perceptual image processing algorithm dataset (PIPAL) [62]. These databases contain various types of distortions, along with varying numbers of reference and distorted images, as summarized in Table II.Since the ground truths of these databases are inconsistent regarding the ranges of their MOS values, a linear normalization function is utilized to preprocess these labels to be within the same interval. In addition, a four-parameter regression function is utilized before calculating the PLCCs and visualizing the prediction scores; this regression equation is formulated as follows:

$$\hat{\gamma} = \frac{\alpha_1 - \alpha_2}{1 + e^{-(\gamma - \alpha_3)/|\alpha_4|}} + \alpha_2, \tag{14}$$

where γ and $\hat{\gamma}$ represent the original input and the output obtained after the regression process, respectively, and $\alpha_{1,\cdots,4}$ are the regression parameters.

B. Performance Comparison

Overall performance evaluation. Performance comparison experiments are conducted on six IQA databases, i.e., LIVE, CSIQ, TID2008, TID2013, KADID-10k and PIPAL, which are divided into training, validation, and testing parts in terms of the reference images. The three components of each dataset

TABLE IV
RESULTS OF A PERFORMANCE COMPARISON CONDUCTED ON THE PIPAL DATASET.

Method	PIPAL Validation [62]	
	PLCC	SRCC
PSNR	0.289	0.255
SSIM [7]	0.409	0.363
MS-SSIM [8]	0.566	0.491
VIF [63]	0.572	0.503
MAD [58]	0.702	0.649
FSIM [30]	0.562	0.468
VSI [64]	0.515	0.450
GMSD [29]	0.656	0.585
NLPD [65]	0.558	0.370
LPIPS-Alex [73]	0.606	0.569
LPIPS-VGG [73]	0.611	0.551
PieAPP [14]	0.696	0.704
DISTS [17]	0.634	0.608
IQT [21]	0.840	0.820
AHIQ [69]	0.845	0.835
UniDASTN [18]	0.831	0.829
GRIDS (ours)	0.838	0.831

Performance comparison between the results obtained by our method and the state-of-the-art methods on the PIPAL dataset. In each column, the best, second-best and third-best results are highlighted in red, green and blue, respectively.

TABLE V
CROSS-DATABASE PERFORMANCE COMPARISON.

Method	LIVE [57]		CSIQ [58]		TID2013 [60]	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
WaDIQaM [39]	0.940	0.947	0.901	0.909	0.834	0.831
PieAPP [14]	0.908	0.919	0.877	0.892	0.859	**0.876**
LPIPS [73]	0.934	0.932	0.896	0.876	0.749	0.670
DISTS [17]	0.954	**0.954**	0.928	0.929	0.855	0.830
GRIDS (ours)	**0.956**	**0.954**	**0.943**	**0.942**	**0.868**	0.861

Cross-database comparison results obtained when training on the entire KADID-10k database.

do not overlap with each other, and the best model determined during an experiment is selected based on the results obtained on the validation set. For the LIVE database, 29 reference images are divided into three sets of $[17, 6, 6]$ images for training, validation, and testing; the CSIQ images are split into $[20, 5, 5]$ images; the TID2008 and TID2013 images are both divided into $[15, 5, 5]$ images; and the KADID-10k database is divided into three sets of $[59, 16, 16]$ images. In particular, we adhere to the official partitioning protocol of the PIPAL dataset, delineating it into designated training, validation, and test subsets. Due to the closure of the competition channel, the ground-truth labels for the test subset are inaccessible. Hence, we rely upon the results obtained from the validation subset for assessing our approach and evaluating all the compared methodologies. The training-free methods are directly tested on each entire database. The experimental results are presented in Table III and Table IV, where the compared methods share a common data partitioning scheme with that used by our approach in the training phase. It can be affirmed that our method achieves commendable performance on the experimental benchmark databases, showing its competitive results in terms of both the PLCC and SRCC. By capturing the complex dependencies and interactions between different regions of an image, the proposed method excels at understanding spatial context information, effectively discerning the influences of semantic content and distortion patterns. Moreover, the incorporation of hierarchical distance measurements enables a more comprehensive and refined image quality analysis. This multiscale approach enables our method to adapt to comprehensive perception variations and integrate information at various levels, providing a holistic assessment of the perceived quality.

Cross-database performance evaluation. We conduct cross-database experiments to evaluate the generalizability of our model in comparison with that of other deep learning-based methods. Given the variations in the sizes of the employed databases, all the compared methods are trained on the comprehensive KADID-10k dataset and subsequently evaluated on distinct databases, including LIVE, CSIQ, and TID2013. The results, as presented in Table V, highlight the exceptional generalization performance of our method, demonstrating its ability to deliver accurate predictions when confronted with previously unseen images. These results emphasize the robustness of the proposed model and its ability to extend its predictive capabilities to diverse datasets.

Evaluation of different distortion subtypes. To further evaluate the performance of our method on different distortion types, we compare the results obtained for traditional distortions and those caused by image processing algorithms, where our method is trained on the entire KADID-10k database for 50 epochs. Figure 5 shows scatter plots of the model prediction results yielded by the representative methods versus the MOSs of the TID2013 database with respect to different distortion subtypes. Our approach exhibits consistent and commendable predictive performance across various distortion subtypes. The ability to produce accurate predictions for different forms of distortion, such as compression artefacts, noise, and blurring, underscores the versatility of the proposed method and its capacity to capture and assess perceptual quality under diverse conditions.

Furthermore, by scrutinizing the specific distortion subtypes, we can identify the strengths and potential areas for improvement exhibited by our method. For instance, the proposed method excels at handling complex image distortions associated with noise and content. Distortions such as high-frequency noise, joint photographic experts group (JPEG)

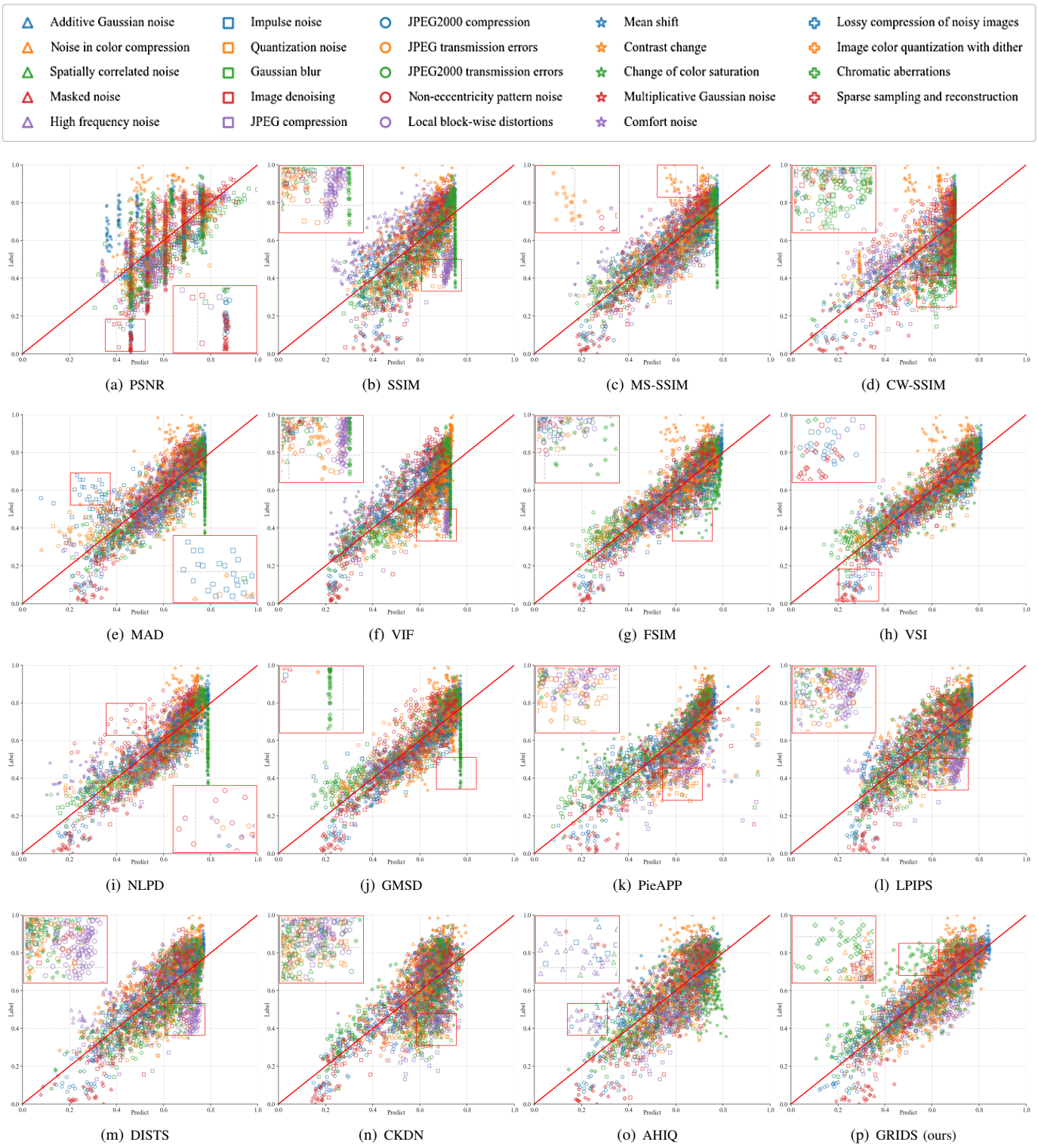


Fig. 5. Scatter plots of the prediction results obtained for the different distortion subtypes in the TID2013 database. Areas with inconsistent predicted scores and densely packed dots are enlarged (as in the red box). Among the compared methods, the values of the lower-better methods are reversed before the regression step.

compression distortion, and colour saturation variations may present challenges for some methods, while our approach effectively addresses these issues. The distribution-based metric aptly reflects the differences between the presence and absence of noise as well as the variations caused by colour saturation changes. The graph representation establishes connections between different content regions, which is particularly beneficial for content-based JPEG compression distortion. However, the performance achieved by our method under chromatic aberration distortion is not satisfactory. This can be attributed

TABLE VI
PERFORMANCE COMPARISON RESULTS OF THE ABLATION EXPERIMENTS.

Method			LIVE [57]		CSIQ [58]		TID2013 [60]	
CC	HD	JSD	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
✔	✔	✗	**0.956**	**0.953**	**0.943**	**0.942**	**0.868**	**0.861**
✔	✗	✔	0.951	0.949	0.935	0.933	0.857	0.852
✗	✔	✗	0.938	0.935	0.923	0.921	0.845	0.839
✗	✗	✔	0.934	0.932	0.917	0.916	0.841	0.833

Ablation experiments are performed after training on KADID-10K. CC designates clique construction. HD and JSD denote the Hellinger distance and Jensen-Shannon divergence, respectively.

TABLE VII
COMPARISON AMONG THE RESULTS OBTAINED WITH DIFFERENT NUMBERS OF NEIGHBOURS.

Neighbours	6	9	12	15
PLCC/SRCC	0.934/0.933	**0.943/0.942**	0.941/0.941	0.936/0.935

Experiments are conducted on the CSIQ database after training on KADID-10K.

to the spatial flexibility of our graph model compared to that of CNNs, which can automatically acquire contextual knowledge about the surrounding space. However, our graph-based model requires more specific training data to achieve optimal performance. This may represent a potential limitation of our approach.

In conclusion, the comprehensive analysis reveals that our method effectively leverages its graph-based architecture to capture the intricate relationships and dependencies within images, enabling it to adapt to various distortion subtypes. The ability to provide consistent and reliable predictions across diverse image degradation scenarios signifies the potential for using the proposed method in practical applications, where images may exhibit complex combinations of distortions.

C. Ablation Study and Parameter Analysis

Impact of clique construction and distribution measurement. We perform ablation experiments to rigorously assess the effectiveness of our graph-based distribution measurement index. Specifically, within our experimental setup, CC describes the application of clique construction operations on a graph, which is a key component of our method. Conversely, the feature maps devoid of clique construction undergo direct comparisons on the original graph. To evaluate the distribution distance between different features, we employ two well-established metrics: the Hellinger distance (HD) and Jensen-Shannon divergence (JSD). We train our model on the KADID-10K database and test it on other databases using different combinations of the abovementioned metrics without modifying the network module. The detailed results are reported in Table VI. Notably, the outcomes consistently underscore the superiority of our graph-based approach, particularly when clique construction is employed. The inclusion of clique construction reveals its significant advantages in terms of capturing intricate semantic and distortion information, enhancing the robustness and effectiveness of our method across various scenarios. These findings reaffirm its generalizability and its potential to significantly enhance distribution-based measurement techniques in image analysis scenarios.

(a) front (b) left (c) rear (d) right (e) top (f) bottom

Fig. 6. Example representations of the six viewing perspectives for PC images.

Number of neighbours. The number of neighbours, denoted as K, is a crucial hyperparameter that shapes the structure of a graph-represented image. This hyperparameter influences the capacity to exchange information and the sizes of the constructed cliques. In our experiments, we conduct trials with K values ranging from 6 to 15, while the other modules and loss functions remain unchanged. The results obtained after training on the KADID-10K dataset and testing on the CSIQ dataset are presented in Table VII. The findings indicate that our method excels at completing image quality tasks when the number of neighbours falls within the range of 9 to 12. An excessive number of neighbours may lead to significant oversmoothing, while overly sparse selections may cause a decline in network performance.

D. Point Cloud Quality Assessment

In this section, we conduct experiments to explore the applicability of our graph-based approach in point cloud quality assessment (PCQA) tasks [79], [81], [83], [85]. With their distinctive attributes and challenges, PC data provide a compelling domain for evaluating our methodology. Importantly, due to the substantial disparities between PC and 2-dimensional (2D) image data, we employ a projection preprocessing step for the PC images. Specifically, we project the data of three-dimensional PC images onto six fixed viewing perspectives, resulting in six two-dimensional images, as shown in Figure 6. These images are resized and subsequently used as inputs for our model. Quality predictions are obtained by assessing the dissimilarity between the reference and distorted images across these multiple projected perspectives. The final predicted PC image quality is computed as the average quality across these six image perspectives.

TABLE VIII
RESULTS OF A PERFORMANCE COMPARISON CONDUCTED AMONG DIFFERENT PCQA METHODS ON THE POINT CLOUD QUALITY DATASETS.

	Method	SJTU-PCQA [75]		WPC [76]	
		PLCC	SRCC	PLCC	SRCC
NR	NIQE [77]	0.376	0.221	0.396	0.389
	BRISQUE [78]	0.224	0.205	0.418	0.378
	3DQA [79]	0.738	0.714	0.651	0.648
FR	p2point-MSE [80]	0.812	0.729	0.485	0.456
	p2point-HAU [80]	0.775	0.716	0.397	0.289
	p2plane-MSE [81]	0.594	0.628	0.270	0.328
	p2plane-HAU [81]	0.687	0.644	0.275	0.283
	PSNR-YUV [82]	0.817	0.795	0.530	0.449
	PCQM [83]	0.885	0.864	0.750	0.743
	GraphSIM [55]	0.845	0.878	0.616	0.583
	PointSSIM [84]	0.714	0.687	0.467	0.454
	IW-SSIM$_p$ [76]	0.795	0.783	0.850	0.848
	GRIDS (ours)	0.909	0.902	0.823	0.821

In each column, the best, second-best and third best-results are highlighted in red, green and blue, respectively.

We perform experiments on two PCQA datasets: the subjective point cloud assessment database (SJTU-PCQA) [75] and the Waterloo point cloud assessment database (WPC) [76]. The SJTU-PCQA dataset is composed of 387 PCs, including 9 reference PCs and 378 distorted PCs. Each reference PC is subjected to 7 diverse distortion types, which are categorized into 6 degradation levels. The WPC dataset comprises 20 reference PCs and 740 distorted PCs. The distortions are produced using 4 unique methods, which include downsampling, Gaussian noise, moving pictures experts group (MPEG)-geometry point cloud compression (MPEG-GPCC), and MPEG-video point cloud compression (MPEG-VPCC). Due to the limited sample size of the dataset, we leverage the version of our model pretrained on the 2D IQA dataset and conduct leave-one-out (LOO) cross-validation according to the reference images within the PCQA dataset for evaluation purposes. In the comparative analysis, we assess the performances of several representative no-reference and full-reference PCQA methods, namely, the natural image quality evaluator (NIQE) [77], the blind/referenceless image spatial quality evaluator (BRISQUE) [78], 3DQA [79], point-to-point (p2point) [80] and point-to-plane (p2plane) [81], which are integrated into MPEG evaluation metrics, PSNR-YUV [82], the point cloud quality metric (PCQM) [83], graph similarity (GraphSIM) [55], PointSSIM [84] and the information content-weighted SSIM (IW-SSIM$_p$) [76]. As shown in Table VIII, the experimental results demonstrate strong performance, signifying the ability of our methodology to perform well in this context. We observe that our graph-based approach effectively captures and evaluates the comprehensive perception characteristics contained within PC data. In particular, our approach exhibits superior performance compared to that of the traditional methods. The robustness and precision of our method highlight its ability to accurately evaluate the quality of PC images.

V. CONCLUSION

In this paper, we presented a hierarchical deep FR-IQA method named GRIDS to measure the perceptual quality distance between distorted and reference images through a comparison of their distribution differences under the graph representation setting. Graph-represented images were automatically generated from the given image content to establish holistic perceptual connections for irregular regions, which were more versatile and effective for the visual perception task. To reflect the perceived image distribution, we decomposed its undirected graph into cliques and then calculated the product of the potential functions for the cliques to obtain the joint probability distribution of the undirected graph. Moreover, we compared the regional distances between the graph distributions of the distorted and reference images at different stages; thus, the distortion distribution measurements derived from different graph model depths were combined to determine the perceived visual quality of the distorted images. Experiments conducted on six benchmark IQA databases and PCQA applications demonstrated the effectiveness of our approach in terms of predicting perceptual image distortions.

REFERENCES

[1] L.-H. Chen, C. G. Bampis, Z. Li, A. Norkin, and A. C. Bovik, "Proxiqa: A proxy approach to perceptual optimization of learned image compression," *IEEE Transactions on Image Processing*, vol. 30, pp. 360–373, 2021. 1

[2] F. Zhou, R. Yao, B. Liu, and G. Qiu, "Visual quality assessment for super-resolved images: Database and method," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3528–3541, 2019. 1

[3] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Comparison of full-reference image quality models for optimization of image processing systems," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1258–1281, 2021. 1

[4] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012. 1

[5] M. Zhou, S. Lang, T. Zhang, X. Liao, Z. Shang, T. Xiang, and B. Fang, "Attentional feature fusion for end-to-end blind image quality assessment," *IEEE Transactions on Broadcasting*, vol. 69, no. 1, pp. 144–152, 2022. 1

[6] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "Generalizable no-reference image quality assessment via deep meta-learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1048–1060, 2022. 1

[7] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. 1, 2, 8, 9, 10

[8] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems and Computers, 2003*, vol. 2, 2003, pp. 1398–1402. 1, 2, 8, 9, 10

[9] Z. Wang and E. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *2005 IEEE International Conference on Acoustics, Speech and Signal Processing.*, vol. 2, 2005, pp. ii/573–ii/576. 1, 2

[10] G.-h. Chen, C.-l. Yang, and S.-l. Xie, "Gradient-based structural similarity for image quality assessment," in *2006 International Conference on Image Processing*, 2006, pp. 2929–2932. 1, 2

[11] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009. 1

[12] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011. 1, 2

[13] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014. 1

[14] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "Pieapp: Perceptual image-error assessment through pairwise preference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 1808–1817. 1, 8, 9, 10

[15] Y. Liu, K. Gu, Y. Zhang, X. Li, G. Zhai, D. Zhao, and W. Gao, "Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 929–943, 2020. 1

[16] W. Kim, A.-D. Nguyen, S. Lee, and A. C. Bovik, "Dynamic receptive field generation for full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4219–4231, 2020. 1, 3

[17] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Image quality assessment: Unifying structure and texture similarity," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 5, pp. 2567–2581, 2020. 1, 8, 10

[18] Z. Tang, Z. Chen, Z. Li, B. Zhong, X. Zhang, and X. Zhang, "Unifying dual-attention and siamese transformer network for full-reference image quality assessment," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 19, no. 6, July 2023. 1, 8, 10

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015, p. 1–14. 1

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2016, pp. 770–778. 1

[21] M. Cheon, S.-J. Yoon, B. Kang, and J. Lee, "Perceptual image quality assessment with transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 2021, pp. 433–442. 1, 8, 10

[22] J. You, "Long short-term convolutional transformer for no-reference video quality assessment," in *Proceedings of the 29th ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2021, p. 2112–2120. 1

[23] Y. Cao, Z. Wan, D. Ren, Z. Yan, and W. Zuo, "Incorporating semi-supervised and positive-unlabeled learning for boosting full reference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2022, pp. 5851–5861. 1, 8, 9

[24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. a. Gelly, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021. 1, 8

[25] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, October 2021, pp. 10 012–10 022. 1

[26] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, June 2016, pp. 2014–2023. 2

[27] G. Li, M. Muller, A. Thabet, and B. Ghanem, "Deepgcns: Can gcns go as deep as cnns?" in *Proceedings of the IEEE/CVF international conference on computer vision*, October 2019, pp. 9267–9276. 2, 5

[28] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, "Vision gnn: An image is worth graph of nodes," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 8291–8303. 2

[29] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014. 2, 8, 9, 10

[30] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011. 2, 8, 9, 10

[31] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "Esim: Edge similarity for screen content image quality assessment," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4818–4831, 2017. 2

[32] W. Sun, Q. Liao, J.-H. Xue, and F. Zhou, "Spsim: A superpixel-based similarity index for full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4232–4244, 2018. 2

[33] Z. Tang, Y. Zheng, K. Gu, K. Liao, W. Wang, and M. Yu, "Full-reference image quality assessment by combining features in spatial and frequency domains," *IEEE Transactions on Broadcasting*, vol. 65, no. 1, pp. 138–151, 2019. 2

[34] S.-H. Bae and M. Kim, "A novel image quality assessment with globally and locally consilient visual quality perception," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2392–2406, 2016. 2, 8, 9

[35] A. Ahar, A. Barri, and P. Schelkens, "From sparse coding significance to perceptual quality: A new approach for image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 879–893, 2018. 2

[36] Y. Fu, H. Zeng, L. Ma, Z. Ni, J. Zhu, and K.-K. Ma, "Screen content image quality assessment using multi-scale difference of gaussian," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2428–2432, 2018. 3

[37] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 1676–1684. 3

[38] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "Deepsim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, 2017, machine Learning and Signal Processing for Big Multimedia Analysis. 3

[39] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2018. 3, 8, 9, 10

[40] L. Wu, J.-Z. Cheng, S. Li, B. Lei, T. Wang, and D. Ni, "Fuiqa: Fetal ultrasound image quality assessment with deep convolutional networks," *IEEE Transactions on Cybernetics*, vol. 47, no. 5, pp. 1336–1349, 2017. 3

[41] K. Ma, Z. Duanmu, and Z. Wang, "Geometric transformation invariant image quality assessment using convolutional neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 6732–6736. 3

[42] S. Ahn, Y. Choi, and K. Yoon, "Deep learning-based distortion sensitivity prediction for full-reference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 2021, pp. 344–353. 3

[43] K. Sim, J. Yang, W. Lu, and X. Gao, "Mad-dls: Mean and deviation of deep and local similarity for image quality assessment," *IEEE Transactions on Multimedia*, vol. 23, pp. 4037–4048, 2021. 3

[44] Y. Zhang, D. M. Chandler, and X. Mou, "Quality assessment of screen content images via convolutional-neural-network-based synthetic/natural segmentation," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5113–5128, 2018. 3

[45] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013. 3

[46] H. Wang, F. Zhang, M. Zhang, J. Leskovec, M. Zhao, W. Li, and Z. Wang, "Knowledge-aware graph neural networks with label smoothness regularization for recommender systems," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 968–977. 3

[47] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph neural networks for social recommendation," in *The world wide web conference*, 2019, pp. 417–426. 3

[48] H. Peng, J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song, and Q. Yang, "Large-scale hierarchical text classification with recursively regularized deep graph-cnn," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 1063–1072. 3

[49] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 7370–7377. 3

[50] M. Xu, P. Fu, B. Liu, and J. Li, "Multi-stream attention-aware graph convolution network for video salient object detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 4183–4197, 2021. 3

[51] M. Meyer, G. Kuschk, and S. Tomforde, "Graph convolutional networks for 3d object detection on radar data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3060–3069. 3

[52] Z. Liang, M. Yang, L. Deng, C. Wang, and B. Wang, "Hierarchical depthwise graph convolutional neural network for 3d semantic segmentation of point clouds," in *2019 International Conference on Robotics and Automation*. IEEE, 2019, pp. 8152–8158. 3

[53] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2020. 3

[54] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual review of neuroscience*, vol. 24, no. 1, pp. 1193–1216, 2001. 3

[55] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 6, pp. 3015–3029, 2020. 8, 13

[56] Z. Shan, Q. Yang, R. Ye, Y. Zhang, Y. Xu, X. Xu, and S. Liu, "Gpa-net:no-reference point cloud quality assessment with multi-task graph convolutional network," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–13, 2023. 8

[57] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006. 9, 10, 12

[58] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010. 8, 9, 10, 12

[59] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009. 9

[60] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015. 9, 10, 12

[61] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted iqa database," in *2019 Eleventh International Conference on Quality of Multimedia Experience*, 2019, pp. 1–3. 9

[62] G. Jinjin, C. Haoming, C. Haoyu, Y. Xiaoxing, J. S. Ren, and D. Chao, "Pipal: A large-scale image quality assessment dataset for perceptual image restoration," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 633–651. 9, 10

[63] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006. 8, 9, 10

[64] L. Zhang, Y. Shen, and H. Li, "Vsi: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014. 8, 9, 10

[65] V. Laparra, J. Balle, A. Berardino, and E. Simoncelii, "Perceptual image quality assessment using a normalized laplacian pyramid," ser. Human Vision and Electronic Imaging 2016, HVEI 2016, 2016, pp. 43–48. 8, 9, 10

[66] H. Lin, V. Hosu, and D. Saupe, "Deepfl-iqa: Weak supervision for deep iqa feature learning," *arXiv preprint arXiv:2001.08113*, 2020. 8, 9

[67] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, July 2017, pp. 1676–1684. 8, 9

[68] S. Seo, S. Ki, and M. Kim, "A novel just-noticeable-difference-based saliency-channel attention residual network for full-reference image quality predictions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2602–2616, 2021. 8, 9

[69] S. Lao, Y. Gong, S. Shi, S. Yang, T. Wu, J. Wang, W. Xia, and Y. Yang, "Attentions help cnns see better: Attention-based hybrid image quality assessment network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, June 2022, pp. 1140–1149. 8, 9, 10

[70] A. Chubarau and J. Clark, "Vtamiq: Transformers for attention modulated image quality assessment," *arXiv preprint arXiv:2110.01655*, 2021. 9

[71] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017. 8

[72] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE transactions on image processing*, vol. 18, no. 11, pp. 2385–2401, 2009. 8

[73] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 586–595. 8, 10

[74] H. Zheng, H. Yang, J. Fu, Z.-J. Zha, and J. Luo, "Learning conditional knowledge distillation for degraded-reference image quality assessment," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 242–10 251. 8

[75] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun, "Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration," *IEEE Transactions on Multimedia*, vol. 23, pp. 3877–3891, 2020. 13

[76] Q. Liu, H. Su, Z. Duanmu, W. Liu, and Z. Wang, "Perceptual quality assessment of colored 3d point clouds," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 8, pp. 3642–3655, 2023. 13

[77] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013. 13

[78] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012. 13

[79] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3d colored point cloud and mesh models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7618–7631, 2022. 12, 13

[80] R. Mekuria, Z. Li, C. Tulvan, and P. Chou, "Evaluation criteria for point cloud compression," *ISO/IEC MPEG*, no. 16332, 2016. 13

[81] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *2017 IEEE International Conference on Image Processing*. IEEE, 2017, pp. 3460–3464. 12, 13

[82] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Applications of Digital Image Processing XLI*, vol. 10752. SPIE, 2018, pp. 174–190. 13

[83] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "Pcqm: A full-reference quality metric for colored 3d point clouds," in *2020 Twelfth International Conference on Quality of Multimedia Experience*, 2020, pp. 1–6. 12, 13

[84] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *2020 IEEE International Conference on Multimedia & Expo Workshops*, 2020, pp. 1–6. 13

[85] Z. Zhang, W. Sun, Y. Zhu, X. Min, W. Wu, Y. Chen, and G. Zhai, "Treating point cloud as moving camera videos: A no-reference quality assessment metric," *arXiv preprint arXiv:2208.14085*, 2022. 12

Wenhao Shen is currently a Ph.D. student of Computer Science in Chongqing University. He received bachelor degree of Computer Science in Chongqing University of Technology. His research interests include perceptual image/video quality assessment and video coding.

Mingliang Zhou rreceived the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2017. He was a Postdoctoral Researcher with the Department of Computer Science, City University of Hong Kong, Hong Kong, China, from September 2017 to September 2019. He was a Postdoctoral Fellow with the State Key Lab of Internet of Things for Smart City, University of Macau, Macau, China, from October 2019 to October 2021. He is currently an Associate Professor with the School of Computer Science, Chongqing University, Chongqing, China. His research interests include image and video coding, perceptual image processing, multimedia signal processing, rate control, multimedia communication, machine learning, and optimization.

Jun Luo received the B.S. and M.S. degrees in mechanical engineering from Henan Polytechnic University, Jiaozuo, China, in 1994 and 1997, respectively, and the Ph.D. degree in mechatronics engineering from the Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai, China, in 2000. He is currently a Professor with the State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, China. His research interests include artificial intelligence, sensing technology, and special robotics.

Zhengguo Li (Fellow, IEEE) received the B.Sci. (applied mathematics) and M.Eng. (automatic control) degrees from Northeastern University, Shenyang, China, in 1992 and 1995, respectively and the Ph.D. degree (automatic control) from Nanyang Technological University, Singapore, in 2001. His research interests include video processing and delivery, computational photography, switched and impulsive control, sensor fusion and physics-driven deep learning. He has co-authored one monograph, more than 200 journal/conference papers including more than 60 IEEE Transaction papers, and eleven granted USA patents, including normative technologies on scalable extension of H.264/AVC and HEVC. Since 2002, he has been actively involved in the development of H.264/AVC and HEVC. He had three informative proposals adopted by the H.264/AVC and three normative proposals adopted by the HEVC. He is currently with the Agency for Science, Technology and Research, Singapore. Zhengguo was the TPC Chair of IEEE IECON 2023 and 2020, special session chair of IEEE ICASSP 2022, Technical Brief Co-Founder of SIGGRAPH Asia, and leading Workshop Chair of IEEE ICME in 2013. He was an Associate Editor for IEEE SIGNAL PROCESSING LETTERS from 2014 to 2016, IEEE TRANSACTIONS ON IMAGE PROCESSING from 2016 to 2020, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2020-2023, APSIPA Transactions on Signal and Information Processing from 2022 to 2025, the Editor of MDPI Sensors from 2022 to 2024, and Senior Area Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING from 2020 to 2024.

Sam Kwong (Fellow, IEEE) received the B.Sc. degree in electrical engineering from The State University of New York, Buffalo, NY, USA, in 1983, the M.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996. He was a Diagnostic Engineer with Control Data Canada, Mississauga, ON, Canada, from 1985 to 1987. He was a Scientific Staff Member with Bell Northern Research Canada, Ottawa, ON, Canada. He is currently a Chair Professor of Computational Intelligence and concurrently as an Associate Vice President (Strategic Research) with Lingnan University, Hong Kong, China. He was elevated to IEEE Fellow in 2014 for his contributions to optimization techniques in cybernetics and video coding. He is the President of the IEEE Systems, Man, and Cybernetics Society from 2021 to 2023. He is also an Associate Editor of several leading IEEE Transactions. He was listed as one of the Top 1% of the World's Most Cited Scientists by Clarivate in 2022.