information

Article

Hybrid No-Reference Quality Assessment for Surveillance Images

Zhongchang Ye, Xin Ye and Zhonghua Zhao *

School of Sensing Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
* Correspondence: zhaozh@sjtu.edu.cn

Abstract: Intelligent video surveillance (IVS) technology is widely used in various security systems. However, quality degradation in surveillance images (SIs) may affect its performance on vision-based tasks, leading to the difficulties in the IVS system extracting valid information from SIs. In this paper, we propose a hybrid no-reference image quality assessment (NR IQA) model for SIs that can help to identify undesired distortions and provide useful guidelines for IVS technology. Specifically, we first extract two main types of quality-aware features: the low-level visual features related to various distortions, and the high-level semantic information, which is extracted by a state-of-the-art (SOTA) vision transformer backbone. Then, we fuse these two kinds of features into the final quality-aware feature vector, which is mapped into the quality index through the feature regression module. Our experimental results on two surveillance content quality databases demonstrate that the proposed model achieves the best performance compared to the SOTA on NR IQA metrics.

Keywords: surveillance image (SI); no-reference image quality assessment (NR IQA); quality-aware features

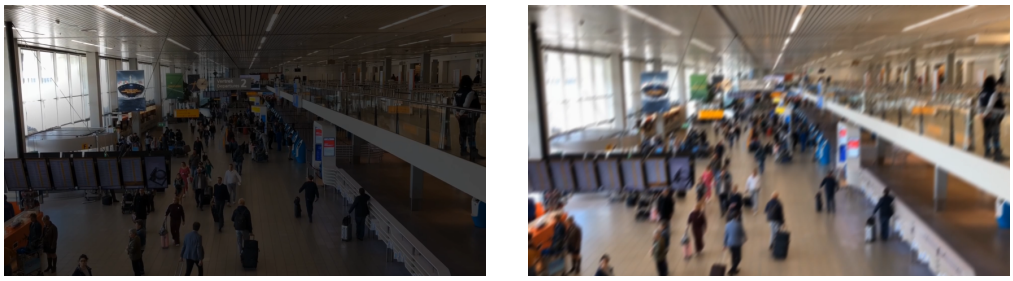
Citation: Ye, Z.; Ye, X.; Zhao, Z. Hybrid No-Reference Quality Assessment for Surveillance Images. *Information* **2022**, *13*, 588. https://doi.org/10.3390/info13120588

Academic Editors: Xin Ning, Yizhang Jiang and Weiwei Cai

Received: 30 October 2022
Accepted: 14 December 2022
Published: 16 December 2022

1. Introduction

With the increasing demand for public security and the rapid development of computer vision technologies and digital products, intelligent video surveillance (IVS) technology has become a hot topic [1]. IVS technology mainly adopts algorithms related to computer vision tasks such as recognition, detection, and tracking in order to understand the content of surveillance videos and automatically perform the task of monitoring or control, which can greatly reduce the burden of human attention [2–5]. Hence, IVS technology has been widely applied in security systems and distributed in various scenarios. However, surveillance images (SIs) usually suffer from different types and degrees of quality degradation in the SI acquisition and transmission process. Specifically, poor physical conditions (smoke, fog, insufficient illumination, etc.), in-capture distortions (noise, blur, etc.), and compression distortions are the main reasons for quality degradations of SIs [6–9]. Distortions in SIs may affect the performance of subsequent high-level tasks, making it difficult for IVS technology to extract valid information from the SIs. As shown in Figure 1, SIs may suffer from uneven illumination or motion blur distortion, leading to difficulties in recognizing objects for both observers and computers. Therefore, it is necessary to consider SI quality assessment (SIQA) in the design of IVS technology. On the one hand, IVS systems can adopt the SIQA method to predict the quality level of the SIs and filter low-quality SIs. On the other hand, IVS systems can employ the SIQA method to detect and identify different types of degradation and apply appropriate quality enhancement processing to improve the quality of the SIs [10]. Both of the two strategies mentioned above can help to improve the performance of IVS systems on vision-based tasks.



(**a**) Uneven illumination (**b**) Motion blur

Figure 1. Examples of quality degradations in SIs: (**a**) the SI suffers from uneven illumination distortion; (**b**) the SI suffers from motion blur distortion.

In the past two decades, IQA has gained popularity in the field of image processing [11]. Depending on whether a human is involved or not, IQA can be divided into subjective IQA and objective IQA [12–14]. Because human eyes are generally the final receiver of the images, subjective IQA is the most reliable way to assess the quality of images. In recent years, many popular IQA databases have been proposed, such as LIVE [15], TID2008 [16], TID2013 [17], CSIQ [18], etc., which are used to train and validate objective IQA methods. Despite having high reliability, subjective IQA methods require lots of time and labour, and as a result are not suitable for real applications. Therefore, objective IQA methods which can automatically predict the quality of images have attracted much attention from researchers and been widely used in various real-world applications [19]. According to the available the reference information, objective IQA can be divided into full-reference IQA (FR IQA), reduced-reference IQA (RR IQA), and no-reference IQA (NR IQA) [20]. FR IQA utilizes whole reference information, while RR IQA adopts partial reference information. The reference signal is not used in NR IQA metrics. In reality, reference images are not available to an IVS systems. Therefore, in this paper we mainly discuss NR IQA methods, as these are more suitable for real applications; however, the absence of a reference image makes them more challenging.

1.1. Related Work

1.1.1. IQA Databases

IQA databases are divided into traditional and emerging databases based on the image content type and underlying application [11]. Traditional databases are generally composed of a few high-quality pristine images and many distorted images, which are corrupted by such typical distortion types as JPEG and JPEG 2000 compression, white noise, blur, etc. LIVE [15] contains 29 reference images and 779 distorted images generated by five common types of distortion. TID2013 [17] consists of 25 pristine images and 3000 distorted images corrupted by 24 distortion types and five distortion levels. CSIQ [18] includes 30 reference images and 866 distorted images generated by six distortion types. These traditional databases sometimes cannot cover the content types and distortion types of certain specific IQA problems. Hence, emerging databases [21–23] have been proposed for specific IQA applications such as 3D images, screen content image, and virtual reality image databases. Because SIs have more complicated content and distortion compared to traditional images, Zhu et al. [7] constructed a surveillance image quality database (SIQD) including 500 in-the-wild SIs with various scenarios, resolutions, and illumination conditions, then performed a study on the subjective quality assessment of these SIs with different degrees of quality. For the surveillance video quality assessment, Beghdadi et al. [8] established the Video Surveillance Quality Assessment Dataset (VSQuAD), which includes 36 reference surveillance videos and 1576 distorted videos generated by nine distortion types.

1.1.2. NR IQA Metrics

Based on the goal of predicting the perception of human vision without any information from the original reference image, many NR IQA metrics have been proposed in recent years [24]. According to distortion type, NR IQA metrics can be divided into general-purpose algorithms and distortion-specific algorithms [11]. General-purpose metrics usually have general quality features designed to describe all types of distortions, while distortion-specific metrics use relevant features designed for a specific IQA problem. General-purpose NR IQA methods can be categorized into three types, namely, natural scene statistics (NSS)-based metrics, learning-based metrics, and human visual system (HVS)-based metrics [25].

The motivation behind NSS-based methods is that high-quality natural scene pictures tend to follow certain statistical properties, and quality degradation can be identified where there is a departure from these statistics. NSS-based methods usually contain three common stages, namely, feature extraction, NSS modeling, and feature regression. Saad et al. [26] designed the BLIINDS (blind image integrity notator using DCT statistics) index based on the NSS of the discrete cosine transformation (DCT) domain. BLIINDS-II [27] adopts the generalized Gaussian distribution (GGD) to model the NSS of the DCT coefficients and then obtains the quality-aware features through the GGD model parameters. BRISQUE [28] (blind image spatial quality evaluator) and DIIVINE [29] apply NSS in the spatial domain to develop their algorithms. GMLF [30] was developed based on the joint statistics of the gradient magnitude (GM) map and the Laplacian of Gaussian (LOG) response.

With the rapid development of machine learning techniques, a large number of learning-based NR IQA metrics have been developed in the last few years [31–33]. CORNIA [34] is based on an unsupervised feature learning framework that uses raw image patches as local descriptors and uses soft-assignment for encoding. Xu et al. [35] developed a NR IQA method based on high-order statistics aggregation (HOSA). Zhang et al. [36] designed a deep bilinear convolutional neural network (CNN)-based NR IQA model for both synthetic and authentic distortions by conceptually modeling them as two-factor variations followed by bilinear pooling. HyperIQA [37] was developed based on a self-adaptive hypernetwork architecture, and introduces a multi-scale local distortion-aware module to capture complex distortions.

The working mechanism of HVS is a high degree of prior knowledge in the design of quality-aware features [38–40]. Zhai et al. [41] developed a psychovisual quality measure based on the free energy principle. Gu et al. [42] designed an NR free energy-based robust metric (NFERM) combining spatial NSS features, free energy-based features, and HVS-inspired features such as structural information and gradient magnitude. Gu et al. [35] proposed a six-step blind metric (SISBLIM) for quality assessment of both singly and multiply distorted images by systematically incorporating the single quality prediction of each emerging distortion type and joint effects of different distortion sources.

If the distortion process is known in advance, distortion-specific NR IQA methods are preferred due to their higher robustness and accuracy [43–45]. JPEG compression, JPEG2000 compression, and blur/noise are the most widely studied distortion types. Based on the observation that pixel values change abruptly across the boundary while remaining unchanged along the whole boundary, Lee et al. [46] designed an NR IQA metric for JPEG images by measuring the strength of blocking artifacts. Sheikh et al. [47] developed an NSS-based metric for JPEG 2000 compression based on the assumption that the compression process can disturb nonlinear dependencies in the natural scenes. Narvekar and Karam [48] adopted a probabilistic model to predict the probability of detecting blur at the image edges, then obtained the blur estimation by pooling the cumulative probability of blur detection (CPBD). To the best of our knowledge, there are few studies on the quality assessment of SIs, and existing general-purpose metrics encounter difficulty when handling the complicated content and distortion types present in SIs. Therefore, there is an urgent need to design an effective NR IQA metric for SIs.

1.2. Contributions

In order to address the SIQA problem, we propose a novel NR IQA model for SIs which is able to predict the quality level or distortion type and level of SIs, helping to improve performance of IVS systems on high-level tasks. The proposed model is composed of three modules, namely, a feature extraction module, feature fusion module, and feature regression module. First, based on the assumption that the human perception of SIs is influenced by both the low-level visual properties and the high-level semantic information, we mainly extract the following two types of quality-aware features: low-level visual features related to various distortion types (noise, blur, and structure damage), and high-level semantic features extracted by the transformer backbone. Second, the feature fusion module concatenates the distortion features and semantic features into a final quality-aware feature representation. Finally, the quality-aware feature representation is mapped into a final quality score or distortion type and level assessment in the feature regression module. Our experimental results show that the proposed NR IQA model outperforms the compared state-of-the-art NR IQA metrics on two surveillance content quality databases.

1.3. Structure

The rest of this study is organized as follows. Section 2 introduces the proposed NR IQA model for SIs in detail. Section 3 mainly presents the experimental results and discussion, including the benchmark databases, experimental setup, IQA competitors, evaluation criteria, performance discussion, statistical tests, and ablation study. Finally, our conclusions are presented in Section 4.

2. Proposed Method

The framework of the proposed method is clearly shown in Figure 2, which includes the feature extraction module, the feature fusion module, and the feature regression module.

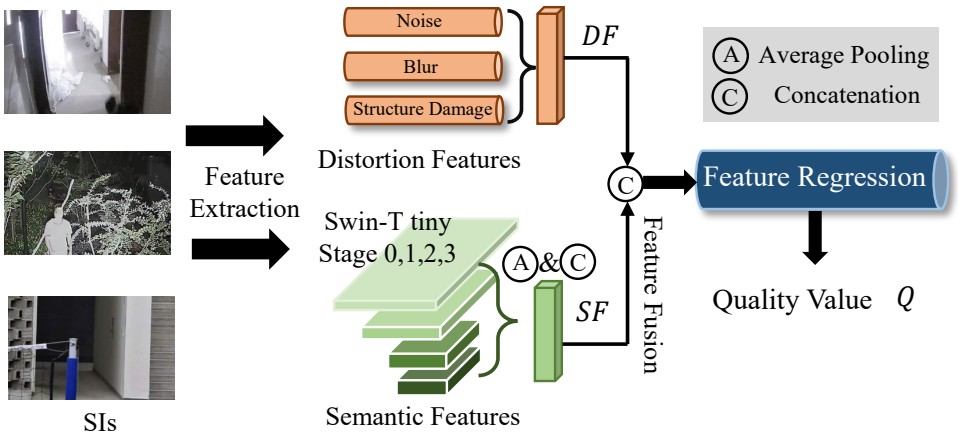


Figure 2. The framework of the proposed method.

2.1. Feature Extraction

SIs can contain various types of distortion, such as noise, blur, structure damage, etc., which inevitably harm the perceived quality. Moreover, the semantic information can influence on human judgment as well [49]. Therefore, in order to fully investigate the information that affects human perception of SIs, we propose to extract features from both the distortion and semantic elements. The distortion features are extracted using the classic image quality descriptors, while the semantic information is collected with the assistance of the high-performance backbone Swin Transformer (ST) [50].

2.1.1. Preliminaries

To better analyze the distortions of SIs, we conduct the local normalization process in advance, which is a common practice in IQA research. Given an SI I, the illumination map can be computed using the maximum of RGB channels

$$L(i,j) = \max_{c\in\{r,g,b\}} I(i,j), \tag{1}$$

where L denotes the illumination map and i and j represent the pixel indexes of SI. Then, the local mean and variance maps can be derived as follows:

$$\begin{aligned} \mu_L(i,j) &= \sum_{k,l} w_{k,l} L(i+k, j+l), \\ \sigma_L(i,j) &= \sqrt{\sum_{k,l} w_{k,l} [L(i+k, j+l) - \mu(i,j)]^2}, \end{aligned} \tag{2}$$

where w is a local Gaussian weighting window, μ_L represents the local mean map, and σ_L represents the local variance map.

2.1.2. Distortion Feature

Noise Estimation: Due to the limitations of camera devices and the insufficient amount of light in dark environments, SIs can be severely degraded by noise distortions. Thus, estimating the level of noise is significant for predicting the quality levels of SIs. Inspired by the task of assessing quality in low-light conditions, we propose using the noise descriptors in [51] to evaluate the level of noise in SIs. Specifically, the noise level is measured in this way by calculating two traditional noise estimation maps through the Gaussian filter and the median filter in order to eliminate the Gaussian noise [52] and salt-and-pepper noise [53], respectively. Then, the noise level can be described by the difference in the images before and after denoising. An example is exhibited in Figure 3. Given a single SI illumination map L, the denoised maps can be derived as follows:

$$M_n = F_n(L) - L, \tag{3}$$

where M_n indicates the noise difference maps, $n \in \{gaussian, median\}$, and F_n represents the denoising function for Gaussian filters (with kernel size 7×7) and median filters (with kernel size 3×3). In common situations, the low-light and flat regions are more easily affected by noise; thus, we compute the final noise level by pooling the noise difference maps in the low-light and flat regions:

$$\begin{aligned} D_n &= \frac{1}{T_R} \sum_{(i,j)\in R} M_n(i,j), \\ R &= \{(i,j) \mid \mu_L(i,j) < E(\mu_L), \sigma_L < E(\sigma_L)\}, \end{aligned} \tag{4}$$

where D_n indicates the estimated noise levels for Gaussian noise and salt-and-pepper noise, $E(\cdot)$ indicates the average operation, T_R denotes the number of pixels in the flat and low-light regions set R, and set R contains all pixels of SIs witgh local mean and variance values smaller than the average local mean and variance.

(**a**) Noisy SI. (**b**) After Gaussian filtering. (**c**) After median filtering.

Figure 3. Examples of noisy images and denoised gray images. The noise in the low-light and flat regions is reduced after Gaussian and median filtering.

Blur Description: Blur is a significant factor in the quality assessment of SIs. Limited by the resolution of camera devices and influenced by the compression of the transmission systems, the texture and details in the SIs may be lost. However, the texture and details are usually vital for identifying the objects and understanding the content of the SIs. Therefore, we propose including the blur features as distortion features. As shown in Figure 4, the gradient features are employed, as they are highly correlated with high-frequency information and have previously been used to describe sharpness [54,55]. Given a single SI illumination map L, we use the Sobel gradient operator to obtain the gradient maps:

$$G_L = \sqrt{(L \otimes S_x)^2 + (L \otimes S_y)^2}, \tag{5}$$

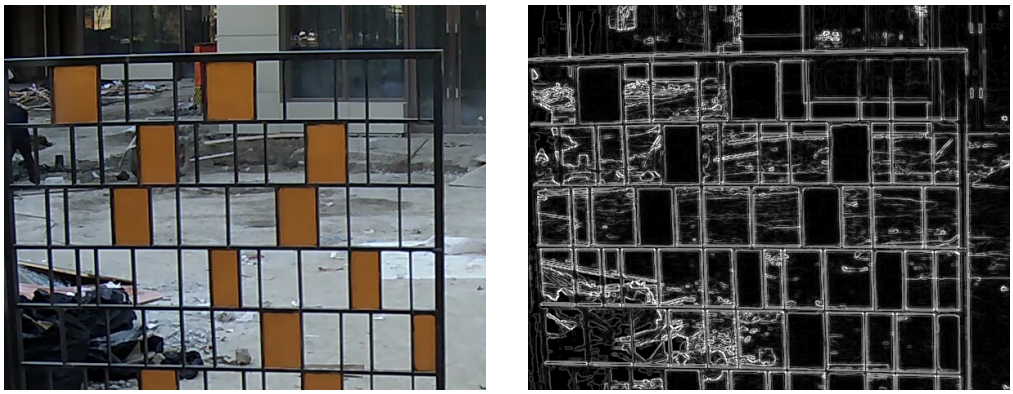
where G_L indicates the gradient magnitude map of the SI illumination map, the operator $\otimes$ represents the convolution operation, and and S_x and S_y are the horizontal and vertical Sobel operators, respectively, which can be described as follows:

$$S_x = \begin{bmatrix} -\frac{1}{4} & 0 & -\frac{1}{4} \\ -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{4} & 0 & -\frac{1}{4} \end{bmatrix}, S_y = \begin{bmatrix} -\frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \end{bmatrix}. \tag{6}$$

With the computed gradient magnitude maps, the blur measurement can be obtained via average pooling:

$$D_b = E(G_L), \tag{7}$$

where D_b is the blur measurement level.



(**a**) Original SI. (**b**) After Sobel operation.

Figure 4. Illustration of an example of original SI and Sobel-operated SI.

Structure Damage: The structure is the outline of the main object in an SI. In this sense, structure damage is caused by low visibility of the major content objects [56]. To quantify the extent of structure damage, we utilize the piecewise smooth image approximation (PSIA) proposed in [57] to generate structure maps:

$$\boldsymbol{Minimize} \quad \boldsymbol{\tau} \tag{8}$$

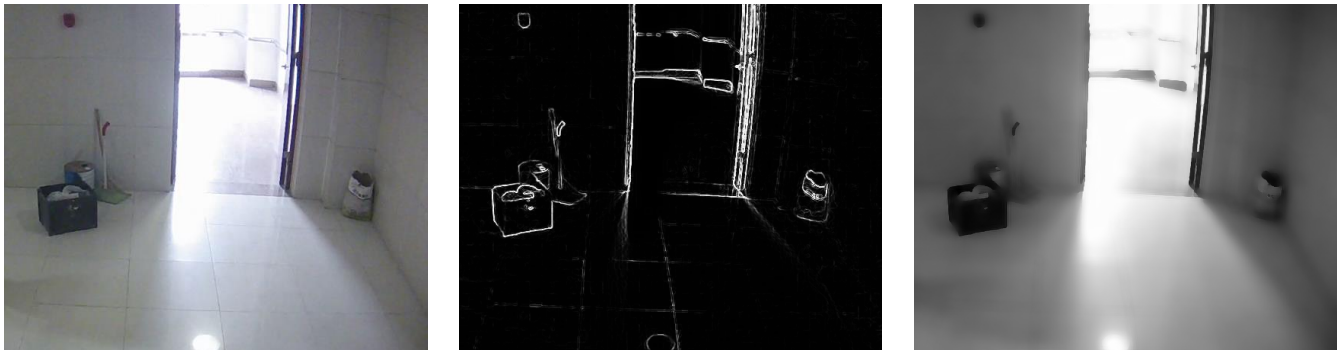
$$\tau = \frac{1}{2}\int_{\Omega}(I - SM)^2 dP + \beta \int_{\Omega \setminus K} |\nabla SM|^2 dP + \alpha \int_{K} d\sigma \tag{9}$$

where SM represents the structure map, Ω is the image domain, K denotes the edge set, $\int_K d\sigma$ represents the total edge length, P indicates the pixel, and the coefficients α and β are positive regularization constants. Figure 5 presents an example of the PSIA results. Similarly, with the computed structure map we can obtain the structure damage descriptor using average pooling:

$$D_s = E(SM), \tag{10}$$

where D_s is the structure damage descriptor.

Summing Up: The process described above results in two noise features D_n ($n \in \{gaussian, median\}$), one blur description feature D_b, and one structure damage feature D_s, which are obtained as the distortion feature vector $DF \in R^{1\times 4}$.



(**a**) Original SI. (**b**) Edge map. (**c**) Structure Map.

Figure 5. Exhibition of piecewise smooth image approximation.

2.1.3. Semantic Feature Extraction

In previous works, it has been proven that the semantic features are highly correlated with quality assessment. Different semantic contents have diverse impacts on human tolerance for different types of distortion [58,59]. For example, humans find blur distortions on flat and texture-free targets such as plain ocean and smooth walls more acceptable. However, blur distortions on objects that are rich in texture, such as rough rocks and complex plants, can be hard to endure. Considering the huge success of the Swin Transformer [50], we use the Swin Transformer-tiny (ST-t) here as the semantic feature extraction backbone. In addition, as visual information is normally perceived hierarchically from low-level to high-level [55], we employ the hierarchical ST-t for feature extraction:

$$\begin{aligned} SF(x) &= \gamma_0(x) \oplus \gamma_1(x) \oplus \gamma_2(x) \oplus \gamma_3(x), \\ \gamma_j(x) &= \mathrm{AP}(F_j(x)), j \in \{0,1,2,3\}, \end{aligned} \tag{11}$$

where $F_j(x)$ denotes the features from the k-th stage, $\mathrm{AP}(\cdot)$ stands for the average pooling operation, $\gamma_k(x)$ denotes the pooled results from the k-th stage, and $\oplus$ indicates the concatenation operation. Then, we can obtain the semantic features $SF \in R^{1\times N_{ST-t}}$, where N_{ST-t} represents the number of output channels of the hierarchical ST-t backbone. Specifically, the dimensions for the feature maps of ST-t's four stages are 784×192, 196×384, 49×768, 49×768. After average pooling, the dimensions turn into 784×1, 196×1, 49×1, and

49×1. After concatenation, the number of the output channels N_{ST-t} of the hierarchical ST-t is 784 + 196 + 49 + 49 = 1078.

2.2. Feature Fusion

In order to actively relate the quality-aware information between the distortion features and semantic features, we first concatenate the features to form one feature vector:

$$\tilde{F} = DF \oplus SF, \tag{12}$$

where $\tilde{F}$ represents the final quality-aware feature vector and $\oplus$ indicates the concatenation operation.

2.3. Feature Regression

There are several tasks in the quality assessment of SIs, including detection of distortion types, identification of the severity level of each detected distortion, and prediction of the overall quality score. In this paper, we design a corresponding feature regression module for each task.

2.3.1. Classification of Distortion Types and Levels

Supposing that the number of distortion types is D_{type} and the number of levels (including the distortion-free level) of each distortion type is D_{level}, we can adopt D_{type} detection branches (DBs) to detect one specific distortion type and estimate the severity level of the corresponding distortion type. Specifically, each DB consists of fully-connected layers containing 128 and D_{level} neurons, respectively. Then, the final quality-aware feature vector $\tilde{F}$ is run through different DBs to obtain the severity level of each distortion type, as follows:

$$P_i = \mathrm{DB_i}(\tilde{F}), i \in \{1, 2, ..., D_{type}\}, \tag{13}$$

where the dimension of the predicted vector P_i is D_{level}, which corresponds to the probability of each severity level for the i-th distortion type. We employ the Cross-Entropy Loss as the loss function for the identification task of each distortion type:

$$L_i = CE(G_i, P_i), \tag{14}$$

where $CE(\cdot)$ refers to the Cross Entropy Loss function and G_i is the ground-truth label of the severity level for the i-th distortion type. Then, we sum the loss functions of all the distortion types to obtain the final loss function:

$$Loss = \sum_{i=1}^{D_{type}} L_i \tag{15}$$

2.3.2. Regression of the Quality Score

With the obtained final quality-aware feature vector $\tilde{F}$, a two-stage fully-connected layer is applied to regress the features into quality scores:

$$Q = FC(\tilde{F}), \tag{16}$$

where $FC(\cdot)$ stands for the fully connected layers and Q represents the regressed quality scores. For the quality assessment tasks, it is necessary to pay attention to the accuracy of the predicted quality levels. Furthermore, the focus should be on the correctness of the quality rankings [49]. Therefore, the loss function employed in this paper includes two parts: the Mean Squared Error (MSE) and the rank error. The MSE loss is employed in order to force the predicted quality values to be close to the quality labels, and can be computed as follows:

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Q_i - Q_i'), \tag{17}$$

where Q_i represents the predicted quality values, Q_i' is the quality label of the SI, and n is the size of the mini-batch. The rank loss has better ability to help the model distinguish the tiny quality difference when the SIs have quite similar quality labels. For this purpose, we employ the differentiable rank function described in [60] to approximate the rank loss:

$$\begin{aligned} L_{rank}^{ij} &= \max\left(0, |Q_i - Q_j| - e(Q_i, Q_j)\cdot\left(Q_i' - Q_j'\right)\right), \\ e(Q_i, Q_j) &= \begin{cases} 1, Q_i \geq Q_j, \\ -1, Q_i < Q_j, \end{cases} \end{aligned} \tag{18}$$

where i and j are the corresponding indexes for two SIs in a mini-batch. The rank loss can be derived as follows:

$$L_{rank} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} L_{rank}^{ij}, \tag{19}$$

Then the loss function can be calculated as the weighted sum of MSE loss and rank loss:

$$Loss = \lambda_1 L_{MSE} + \lambda_2 L_{rank} \tag{20}$$

where λ_1 and λ_2 are used to define the weight of the MSE loss and the rank loss, respectively.

3. Experiment

3.1. Benchmark Databases

We mainly validated our methods on the SI Quality Database (SIQD) [7] and Video Surveillance Quality Assessment Database (VSQuAD) [8]. The SIQD database contains 500 in-the-wild SIs that are diverse in termss of both content and distortions. The main objects in the SIs in the SIQD database include humans and vehicles, and the database covers a wide resolution range, from 352 × 288 to 1920 × 1080. The VSQuAD database contains 964 single-distortion-affected and 612 multiple-distortion-affected surveillance videos (SVs) generated from 36 reference SVs. The distortions include defocus blur, haze, low-light conditions, motion blur, rain, smoke, uneven illumination, and compression artifacts. Each SV lasts for 10 s. Because we propose an IQA method for SIs, we extract ten frames of each SV (one frame for each second) as the representative SIs for each SV. Thus, the extracted SIs have the same distortion types and levels labels as the source SV.

Additionally, we conducted a subjective experiment to gather the quality labels using the SIQD database. Several human participants were invited to judge the quality of the SIs in a well-controlled environment, and their mean opinion scores were recorded as the ground truth for the SIs. For the VSQuAD database, distortions were manually introduced to the surveillance videos, and the type and strength levels of the added distortions were recorded for use as the ground truth.

3.2. Experimental Setup

The employed hierarchical ST-t [50] backbone was initialized with the weights pre-trained on the ImageNet database [61] for semantic feature extraction. The SIs were first resized to the resolution of 256 × 256 and then randomly cropped into patches with the resolution of 224 × 224 as the inputs. The Adam optimizer [62] was utilized, with the initial learning rate set as 1×10^{-4} . The learning rate decays with a ratio of 0.95 every five epochs. The default number of the training epochs was set as 50. If the training loss did not decrease for ten epochs, the training process was ended. Furthermore, we employed the five-fold cross validation strategy. We split the SIQD database into five groups, with each group containing 100 SIs. For each unique group, we trained the model on the left four groups and used the unique group as testing sets. This process was repeated five times to ensure that each group was taken as the testing set only once. Then, the average

performance was recorded as the final performance for the model. A similar five-fold cross validation strategy was conducted on the VSQuAD database.

IQA Competitors

To fully validate the effectiveness of the proposed method, several mainstream IQA methods are selected for comparison, be categorized into two types:

- Hand-crafted methods: BLIINDS-II [27], BRISQUE [28], CORNIA [34], DIIVINE [29], GMLF [30], HOSA [35], SISBLIM [35], and NFERM [42].
- Deep learning-based methods: SFA [63], DBCNN [36], and HyperIQA [37].

It is worth mentioning here that the compared methods were all retrained using the default experimental setup.

3.3. Evaluation Criteria

3.3.1. Classification of the Distortion Types and Levels

To assess the detection of distortion types and identification of the severity level of each detected distortion, we utilize the Accuracy and F1 score to evaluate the predictive performance of different quality assessment metrics. Specifically, the following four evaluation metrics are used:

- $Accu_{type}$: The ratio of correctly predicted observations to the total observations for distortion detection.
- $F1_{type}$: The weighted average of Precision and Recall for distortion detection.
- $Accu_{both}$: The ratio of correctly predicted observations to total observations for distortion detection with severity level identification.
- $F1_{both}$: The weighted average of Precision and Recall for the distortion detection with severity level identification.

3.3.2. Regression of the Quality Score

Here, four criteria are used to evaluate the performance of the quality assessment models: the Spearman Rank Order Correlation Coefficient (SRCC), the Pearson Linear Correlation Coefficient (PLCC), the Kendall Rank Correlation Coefficient (KRCC), and the Root Mean Squared Error (RMSE). These four statistical indexes describe different aspects for evaluating the performance of IQA models. To be more specific, SRCC and KRCC both reflect the prediction monotonicity, while PLCC and RMSE reflect the prediction linearity and prediction accuracy, respectively. The calculation equations are as follows:

- Spearman rank order correlation coefficient (SRCC):

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \tag{21}$$

  where d_i represents the difference between the i-th images's ranks in subjective evaluations and predicted scores, while N is the number of testing images. SRCC is used to measure the prediction monotonicity. The value of SRCC is between 0 and 1. The larger the value, the better the result predicted by the model.
- Pearson linear correlation coefficient (PLCC):

$$PLCC = \frac{\sum_{i=1}^{N}(p_i - \overline{p})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(p_i - \overline{p})^2 (s_i - \overline{s})^2}}, \tag{22}$$

  where s_i and p_i represent the i-th image's subjective score and predicted score, while $\overline{s}$ and $\overline{p}$ are the mean of all s_i and p_i. PLCC can be used to estimate the linearity and consistency of prediction. The value of PLCC is between 0 and 1, with larger values being better.

- Kendall rank order correlation coefficient (KRCC):

$$KRCC = \frac{N_c - N_d}{0.5(N-1)N}, \tag{23}$$

where N_c and N_d represent the numbers of concordant and discordant pairs in the testing data. Similar to SRCC, KRCC can be used to measure the monotonicity. The value of KRCC is between 0 and 1, with larger values being better.

- Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - s_i)^2}, \tag{24}$$

RMSE is used to evaluate prediction accuracy. The RMSE value is a positive number; a smaller the value indicates higher accuracy of the model.

Before computing the criteria values, we utilize a five-parameter logistic regression function to fit the predicted scores to the scale of the quality labels:

$$\hat{y} = \beta_1 \left(0.5 - \frac{1}{1 + e^{\beta_2 (y - \beta_3)}}\right) + \beta_4 y + \beta_5 \tag{25}$$

where $\{\beta_i \mid i = 1, 2, \ldots, 5\}$ are the parameters to be fitted, y represents the predicted scores, and $\hat{y}$ represents the mapped scores.

3.4. Performance Discussion

The experimental performance results on the SIQD and the VSQuAD databases are clearly shown in Tables 1 and 2, from which we can draw several interesting conclusions: (a) the deep-learning based methods achieve much better performance than the hand-crafted based methods, indicating that the semantic information extracted by the CNN or vision transformer backbone is very important for the quality prediction of the SIs; (b) the proposed NR IQA method performs the best on both the SIQA database and VSQuAD dataset compared with other NR IQA metrics, which demonstrates the effectiveness of the proposed NR IQA method for the SIs; (c) the proposed model outperforms all the compared deep-learning based methods, indicating that the low-level visual features related to distortions in the SIs serve as a vital complement to the deep features for the quality assessment of SIs, from which it can be concluded that it is necessary to specifically design relevant features for specific IQA problems; (d) the task of identifying the severity level of each distortion is relatively difficult compared to the task of detecting the distortion type, which demonstrates that quality assessment models are less sensitive to the distortion level.

Table 1. Performance results on the SIQD database.

Type	Method	SRCC	PLCC	KRCC	RMSE
Hand-crafted	BLIINDS-II	0.1584	0.2059	0.0946	0.9030
	BRISQUE	0.3051	0.3256	0.2497	0.8726
	CORNIA	0.5476	0.5641	0.4732	0.7619
	DIIVINE	0.0223	0.2178	0.0132	0.9007
	GMLF	0.0740	0.2058	0.0533	0.9030
	HOSA	0.2871	0.3273	0.2064	0.8720
	SISBLIM	0.4206	0.5488	0.3612	0.7714
	NFERM	0.2576	0.3925	0.2167	0.8488
Deep-learning	SFA	0.8702	0.8741	0.7123	0.4153
	DBCNN	0.8727	0.8785	0.7196	0.4033
	HyperIQA	0.8631	0.8687	0.6946	0.4478
	Proposed	**0.8986**	**0.9103**	**0.7276**	**0.3864**

Table 2. Performance results on the VSQuAD database.

Type	Method	$Accu_{type}$	$F1_{both}$	$Accu_{both}$	$F1_{both}$
Hand-crafted	BLIINDS-II	0.311	0.569	0.051	0.088
	BRISQUE	0.368	0.603	0.076	0.127
	CORNIA	0.540	0.642	0.371	0.422
	DIIVINE	0.270	0.432	0.041	0.067
	GMLF	0.289	0.411	0.048	0.077
	HOSA	0.378	0.615	0.086	0.167
	SISBLIM	0.524	0.634	0.343	0.396
	NFERM	0.603	0.723	0.413	0.487
Deep-learning	SFA	0.762	0.892	0.622	0.672
	DBCNN	0.778	0.903	0.638	0.690
	HyperIQA	0.794	0.911	0.654	0.716
	Proposed	**0.852**	**0.946**	**0.708**	**0.817**

3.5. Statistical Test

To further validate the effectiveness of the proposed method, we carried out statistical significance tests following the procedure suggested in [64]. In this subsection, these statistical tests are used to compare the relations between the predicted results and the subjective labels. The null hypothesis of the t-test is that the residuals of two quality metrics derived from the same distribution are statistically indistinguishable with a 95% confidence. The statistical significance test results are shown in Figure 6. From the figure, it can be seen that the proposed method is significantly superior to nine compared methods on the SIQD database and eleven compared methods on the VSQuAD database, indicating that the proposed method has better ability to detect and evaluate distortions in SIs.

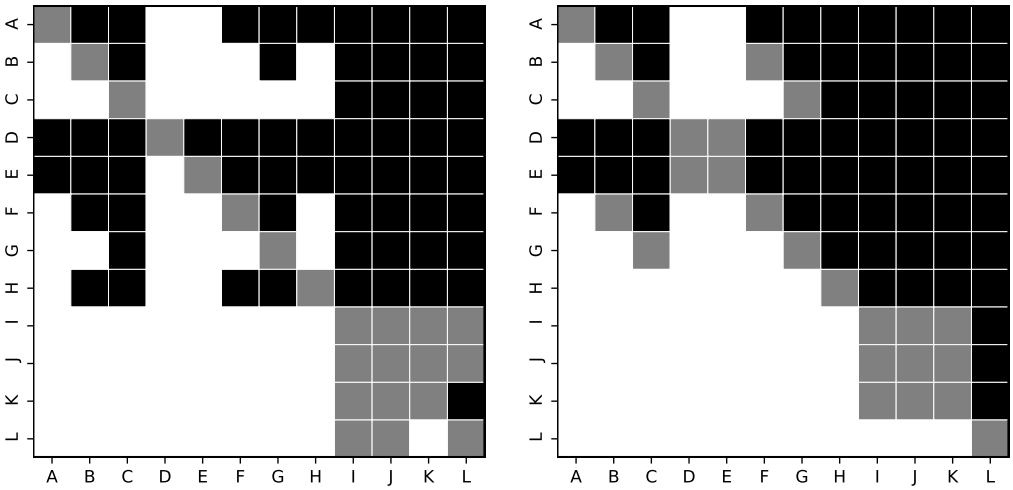


Figure 6. Statistical test results of the proposed method and compared methods on the SIQD and VSQuAD databases: (**a**) statistical test results on the SIQD database and (**b**) results on the VSQuAD database. Black/white blocks mean that the method in that row is statistically worse/better than one in the corresponding column. A gray block means that the method in the row and that in the column are statistically indistinguishable. The methods denoted by A-L are in the same order as in Tables 1 and 2.

3.6. Ablation Study

To further investigate the respective contributions of different types of features, we performed an ablation experiment to compare the distortion features, semantic features and hybrid (distortion and semantic) features.

The results of the ablation experiment are listed in Tables 3 and 4. First, it can be seen that the hybrid features perform better than either the distortion features or the semantic

features alone. Second, the contribution of the distortion features is inferior to that of the semantic features, meaning that the semantic features are more important in quality assessment of SIs. Finally, the semantic features perform worse than all the compared deep learning NR IQA metrics, which may be explained by the resize operation resulting in the loss of texture information, which could in turn affect the performance.

Table 3. Ablation study results on the SIQD database; DF represents distortion features and SF indicates semantic features. The default experimental setup and quality regression mechanism are maintained.

Feature	SRCC	PLCC	KRCC	RMSE
DF	0.6143	0.6268	0.5293	0.6923
SF	0.7738	0.7910	0.6104	0.5312
DF+SF	**0.8986**	**0.9103**	**0.7276**	**0.3864**

Table 4. Ablation study results on the VSQuAD database; DF represents distortion features and SF indicates semantic features. The default experimental setup and quality regression mechanism are maintained.

Feature	$Accu_{type}$	$F1_{both}$	$Accu_{both}$	$F1_{both}$
DF	0.625	0.697	0.441	0.501
SF	0.667	0.757	0.473	0.564
DF+SF	**0.852**	**0.946**	**0.708**	**0.817**

4. Conclusions

To tackle the challenge of SIQA and provide more useful guidelines for surveillance systems, in this paper we propose a hybrid no-reference image quality assessment method. The features are mainly extracted from the distortion and semantic aspects. Specifically, the distortion features are extracted using the noise, blur, and structure hand-crafted descriptors. We employ Swin Transformer-tiny as the backbone for semantic feature extraction, in light of its great success as a vision transformer. Afterwards, the hybrid features are concatenated and regressed into quality values with the assistance of fully-connected layers. The proposed method is validated on the SI Quality Database (SIQD) and the Video Surveillance Quality Assessment Database (VSQuAD). Finally, we evaluate several similar methods and compare them to our proposed method by assessing the correlation between their predicted scores and quality labels and by measuring their accuracy when predicting distortion types and levels. From the experimental results, we find that the proposed method outperforms all the compared methods, revealing its strong ability to solve the SIQA problem.

Author Contributions: Conceptualization, Z.Y. and X.Y.; methodology, Z.Y.; software, Z.Y.; validation, Z.Y., X.Y. and Z.Z.; formal analysis, Z.Y.; investigation, Z.Y.; resources, Z.Y.; data curation, Z.Y.; writing—original draft preparation, Z.Y.; writing—review and editing, Z.Y.; visualization, Z.Y.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IVS	Intelligent Video Surveillance
IQA	Image Quality Assessment
FR IQA	Full-reference IQA
RR IQA	Reduced-reference IQA
NR IQA	No-reference IQA
SI	Surveillance Image
SIQA	Surveillance Image Quality Assessment
SV	Surveillance Video
SIQD	Surveillance Image Quality Database
VSQuAD	Video Surveillance Quality Assessment Database
ST	Swin Transformer
ST-t	Swin Transformer-tiny
DF	Distortion Features
SF	Semantic Features
MSE	Mean Squared Error
SRCC	Spearman Rank Order Correlation Coefficient
PLCC	Pearson Linear Correlation Coefficient
KRCC	Kendall Rank Correlation Coefficient
RMSE	Root Mean Squared Error

References

1. Gonzalez-Cepeda, J.; Ramajo, A.; Armingol, J.M. Intelligent Video Surveillance Systems for Vehicle Identification Based on Multinet Architecture. *Information* **2022**, *13*, 325. https://doi.org/10.3390/info13070325.
2. Sreenu, G.; Durai, M.S. Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *J. Big Data* **2019**, *6*, 1–27.
3. Muller-Schneiders, S.; Jager, T.; Loos, H.S.; Niem, W. Performance evaluation of a real time video surveillance system. In Proceedings of the 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing, China, 15–16 October 2005; pp. 137–143.
4. Aqqa, M.; Mantini, P.; Shah, S.K. Understanding How Video Quality Affects Object Detection Algorithms. In Proceedings of the VISIGRAPP (5: VISAPP), Prague, Czech Republic, 25–27 February 2019; pp. 96–104.
5. Held, C.; Krumm, J.; Markel, P.; Schenke, R.P. Intelligent video surveillance. *Computer* **2012**, *45*, 83–84.
6. Leszczuk, M.; Romaniak, P.; Janowski, L. Quality assessment in video surveillance. In *Recent Developments in Video Surveillance*; IntechOpen: London, UK, 2012.
7. Zhu, W.; Zhai, G.; Yao, C.; Yang, X. SIQD: Surveillance Image Quality Database and Performance Evaluation for Objective Algorithms. In Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 9–12 December 2018; pp. 1–4.
8. Beghdadi, A.; Qureshi, M.A.; Dakkar, B.E.; Gillani, H.H.; Khan, Z.A.; Kaaniche, M.; Ullah, M.; Cheikh, F.A. A New Video Quality Assessment Dataset for Video Surveillance Applications. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 1521–1525.
9. Bezzine, I.; Khan, Z.A.; Beghdadi, A.; Al-Maadeed, N.; Kaaniche, M.; Al-Maadeed, S.; Bouridane, A.; Cheikh, F.A. Video Quality Assessment Dataset for Smart Public Security Systems. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020; pp. 1–5.
10. Zhang, Z.; Lu, W.; Sun, W.; Min, X.; Wang, T.; Zhai, G. Surveillance Video Quality Assessment Based on Quality Related Retraining. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 4278–4282.
11. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. *Sci. China Inf. Sci.* **2020**, *63*, 1–52.
12. Zhai, G.; Sun, W.; Min, X.; Zhou, J. Perceptual quality assessment of low-light image enhancement. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2021**, *17*, 1–24.
13. Mohammadi, P.; Ebrahimi-Moghadam, A.; Shirani, S. Subjective and objective quality assessment of image: A survey. *arXiv* **2014**, arXiv:1406.7799.
14. Golestaneh, S.A.; Dadsetan, S.; Kitani, K.M. No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 3209–3218.
15. Sheikh, H. LIVE Image Quality Assessment Database Release 2. 2005. Available online: http://live.ece.utexas.edu/research/quality (accessed on 20 October 2022)

16. Ponomarenko, N.; Lukin, V.; Zelensky, A.; Egiazarian, K.; Carli, M.; Battisti, F. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radioelectron.* **2009**, *10*, 30–45.
17. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Commun.* **2015**, *30*, 57–77.
18. Larson, E.C.D. Consumer Subjective Image Quality Database. 2009. Available online: http://visionokstateedu/csiq/ (accessed on 20 October 2022).
19. Wang, Z. Applications of objective image quality assessment methods [applications corner]. *IEEE Signal Process. Mag.* **2011**, *28*, 137–142.
20. Wang, L. A survey on IQA. *arXiv* **2021**, arXiv:2109.00347.
21. Benoit, A.; Le Callet, P.; Campisi, P.; Cousseau, R. Quality assessment of stereoscopic images. *EURASIP J. Image Video Process.* **2009**, *2008*, 1–13.
22. Yang, H.; Fang, Y.; Lin, W.; Wang, Z. Subjective quality assessment of screen content images. In Proceedings of the 2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX), Singapore, 18–20 September 2014; pp. 257–262.
23. Ma, K.; Zeng, K.; Wang, Z. Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* **2015**, *24*, 3345–3356.
24. Kim, J.; Lee, S. Deep blind image quality assessment by employing FR-IQA. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3180–3184.
25. Ye, P.; Doermann, D. No-reference image quality assessment using visual codebooks. *IEEE Trans. Image Process.* **2012**, *21*, 3129–3138.
26. Saad, M.A.; Bovik, A.C.; Charrier, C. A DCT statistics-based blind image quality index. *IEEE Signal Process. Lett.* **2010**, *17*, 583–586.
27. Saad, M.A.; Bovik, A.C.; Charrier, C. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.* **2012**, *21*, 3339–3352.
28. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708.
29. Moorthy, A.K.; Bovik, A.C. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Trans. Image Process.* **2011**, *20*, 3350–3364.
30. Xue, W.; Mou, X.; Zhang, L.; Bovik, A.C.; Feng, X. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Trans. Image Process.* **2014**, *23*, 4850–4862.
31. Lu, W.; Sun, W.; Zhu, W.; Min, X.; Zhang, Z.; Wang, T.; Zhai, G. A cnn-based quality assessment method for pseudo 4k contents. In *Proceedings of the International Forum on Digital TV and Wireless Multimedia Communications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 164–176.
32. Wang, T.; Sun, W.; Min, X.; Lu, W.; Zhang, Z.; Zhai, G. A Multi-dimensional Aesthetic Quality Assessment Model for Mobile Game Images. In Proceedings of the 2021 International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany, 5–8 December 2021; pp. 1–5.
33. Sun, W.; Min, X.; Zhai, G.; Gu, K.; Duan, H.; Ma, S. MC360IQA: A multi-channel CNN for blind 360-degree image quality assessment. *IEEE J. Sel. Top. Signal Process.* **2019**, *14*, 64–77.
34. Ye, P.; Kumar, J.; Kang, L.; Doermann, D. Unsupervised feature learning framework for no-reference image quality assessment. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1098–1105.
35. Xu, J.; Ye, P.; Li, Q.; Du, H.; Liu, Y.; Doermann, D. Blind image quality assessment based on high order statistics aggregation. *IEEE Trans. Image Process.* **2016**, *25*, 4444–4457.
36. Zhang, W.; Ma, K.; Yan, J.; Deng, D.; Wang, Z. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *30*, 36–47.
37. Su, S.; Yan, Q.; Zhu, Y.; Zhang, C.; Ge, X.; Sun, J.; Zhang, Y. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3667–3676.
38. Ma, J.; Wu, J.; Li, L.; Dong, W.; Xie, X.; Shi, G.; Lin, W. Blind image quality assessment with active inference. *IEEE Trans. Image Process.* **2021**, *30*, 3650–3663.
39. Gao, X.; Lu, W.; Tao, D.; Li, X. Image quality assessment and human visual system. In Proceedings of the Visual Communications and Image Processing, Huangshan, China, 11–14 July 2010; Volume 7744, pp. 316–325.
40. Zhai, G.; Min, X.; Liu, N. Free-energy principle inspired visual quality assessment: An overview. *Digit. Signal Process.* **2019**, *91*, 11–20.
41. Zhai, G.; Wu, X.; Yang, X.; Lin, W.; Zhang, W. A psychovisual quality metric in free-energy principle. *IEEE Trans. Image Process.* **2011**, *21*, 41–52.
42. Gu, K.; Zhai, G.; Yang, X.; Zhang, W. Using free energy principle for blind image quality assessment. *IEEE Trans. Multimed.* **2014**, *17*, 50–63.
43. Wang, H.; Fu, J.; Lin, W.; Hu, S.; Kuo, C.C.J.; Zuo, L. Image quality assessment based on local linear information and distortion-specific compensation. *IEEE Trans. Image Process.* **2016**, *26*, 915–926.

44. Sadbhawna; Jakhetiya, V.; Mumtaz, D.; Jaiswal, S.P. Distortion specific contrast based no-reference quality assessment of DIBR-synthesized views. In Proceedings of the 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), Tampere, Finland, 21–24 September 2020; pp. 1–5.
45. Yan, C.; Teng, T.; Liu, Y.; Zhang, Y.; Wang, H.; Ji, X. Precise no-reference image quality evaluation based on distortion identification. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2021**, *17*, 1–21.
46. Lee, S.; Park, S.J. A new image quality assessment method to detect and measure strength of blocking artifacts. *Signal Process. Image Commun.* **2012**, *27*, 31–38.
47. Sheikh, H.R.; Bovik, A.C.; Cormack, L. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Trans. Image Process.* **2005**, *14*, 1918–1927.
48. Narvekar, N.D.; Karam, L.J. A no-reference image blur metric based on the cumulative probability of blur detection (CPBD). *IEEE Trans. Image Process.* **2011**, *20*, 2678–2683.
49. Zhang, Z.; Sun, W.; Min, X.; Zhou, Q.; He, J.; Wang, Q.; Zhai, G. MM-PCQA: Multi-Modal Learning for No-reference Point Cloud Quality Assessment. *arXiv* **2022**, arXiv:2209.00244.
50. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF CVPR, Montreal, QC, Canada, 10-17 October 2021; pp. 10012–10022.
51. Zhang, Z.; Sun, W.; Min, X.; Zhu, W.; Wang, T.; Lu, W.; Zhai, G. A no-reference evaluation metric for low-light image enhancement. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6.
52. Chowdhury, D.; Das, S.K.; Nandy, S.; Chakraborty, A.; Goswami, R.; Chakraborty, A. An Atomic Technique For Removal Of Gaussian Noise From A Noisy Gray Scale Image Using LowPass-Convoluted Gaussian Filter. In Proceedings of the International Conference on Opto-Electronics and Applied Optics, Kolkata, India, 18–20 March 2019; pp. 1–6.
53. Chang, C.; Hsiao, J.; Hsieh, C. An Adaptive Median Filter for Image Denoising. In Proceedings of the International Symposium on Intelligent Information Technology Application, Shanghai, China, 20–22 December 2008; Volume 2, pp. 346–350.
54. Zhang, Z.; Sun, W.; Min, X.; Wang, T.; Lu, W.; Zhai, G. A Full-Reference Quality Assessment Metric for Fine-Grained Compressed Images. In Proceedings of the 2021 International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany, 5–8 December 2021; pp. 1–4.
55. Zhang, Z.; Sun, W.; Wu, W.; Chen, Y.; Min, X.; Zhai, G. Perceptual Quality Assessment for Fine-Grained Compressed Images. *arXiv* **2022**, arXiv:2206.03862.
56. Zhang, Z.; Sun, W.; Min, X.; Zhu, W.; Wang, T.; Lu, W.; Zhai, G. A No-Reference Deep Learning Quality Assessment Method for Super-resolution Images Based on Frequency Maps. *arXiv* **2022**, arXiv:2206.04289.
57. Bar, L.; Sochen, N.; Kiryati, N. Semi-blind image restoration via Mumford-Shah regularization. *IEEE Trans. Image Process.* **2006**, *15*, 483–493. https://doi.org/10.1109/TIP.2005.863120.
58. Dodge, S.; Karam, L. Understanding how image quality affects deep neural networks. In Proceedings of the IEEE QoMEX, Lisbon, Portugal, 6–8 June 2016; pp. 1–6.
59. Li, D.; Jiang, T.; Jiang, M. Quality assessment of in-the-wild videos. In Proceedings of the ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2351–2359.
60. Sun, W.; Min, X.; Lu, W.; Zhai, G. A Deep Learning based No-reference Quality Assessment Model for UGC Videos. *arXiv* **2022**, arXiv:2204.14047.
61. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In Proceedings of the IEEE/CVF CVPR, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
62. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference for Learning Representations, Diego, CA, USA, 7–9 May 2015; p. 13.
63. Li, D.; Jiang, T.; Lin, W.; Jiang, M. Which has better visual quality: The clear blue sky or a blurry animal? *IEEE Trans. Multimed.* **2018**, *21*, 1221–1234.
64. Sheikh, H.; Sabir, M.; Bovik, A. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. https://doi.org/10.1109/TIP.2006.881959.