Image Manipulation Quality Assessment

Xinbo Wu, Jianxun Lou, Yingying Wu, Wanan Liu, Paul L. Rosin, Gualtiero Colombo,
Stuart Allen, Roger Whitaker and Hantao Liu

***Abstract*—Image quality assessment (IQA) and its computational models play a vital role in modern computer vision applications. Research has traditionally focused on signal distortions arising during image compression and transmission, and their impact on perceived image quality. However, little attention is paid to image manipulation that alters an image using various filters. With the prevalence of image manipulation in real-life scenarios, it is critical to understand how humans perceive filter-altered images and to develop reliable IQA models capable of automatically assessing the quality of filtered images. In this paper, we build a new IQA database for filter-altered images, comprised of 360 images manipulated by various filters. To ensure the subjective IQA faithfully reflects human visual perception, we conduct a fully-controlled psychovisual experiment. Building upon the ground truth, we propose an innovative deep learning-based no-reference IQA (NR-IQA) model named IMQA that can accurately predict the perceived quality of filter-altered images. This model involves constructing an image filtering-aware module to learn discriminatory features for filter-altered images; and fuses these features with the representations generated by an image quality-aware module. Experimental results demonstrate the superior performance of the proposed IMQA model.**

***Index Terms*—Image quality assessment, perception, filter-altered, image manipulation, deep learning**

I. Introduction

Image quality assessment (IQA) involves understanding how humans perceive image quality and use the understanding to enabling computers to accurately make such assessments. As such, IQA algorithms form the foundation for evaluating, monitoring, and optimising modern computer vision and imaging systems [1]. Recently, the widespread use of mobile devices and the persistent evolution of digital media have made IQA an increasingly important aspect aligned to visual experience, covering various practical applications such as medical image analysis [2]–[4], mobile computing [5]–[7], and image and video coding [8], [9].

Significant progress has been made in IQA research with advancements in both subjective and objective methodologies [10]. In the literature, IQA methods have primarily focused on signal distortions arising during image compression and transmission, where the quality assessment is driven by the perceptual impact of these distortions on viewers' preferences. Because of this research focus, existing IQA databases [11]–[18] are mainly constructed for synthetic and/or authentic image distortions. Nowadays, filter-altered images have become ubiquitous across various social media platforms. However, there remains a distinction in the quality assessment of these filter-altered images. Previous research [19] indicates that quality perception of filter-altered images may differ from that of distorted images, and consequently the IQA-related visual representations of these two types of images may differ when concerning the development of computational IQA models [20]. Therefore, it is crucial to build dedicated IQA databases for filter-altered images. This is needed to facilitate our understanding of visual perception of viewers, as well as the development of computational IQA algorithms that can reliably assess perceived quality of filter-altered images. Also, existing IQA models, including both traditional models [11], [15], [21]–[23] based on handcrafted features and deep learning-based models [24]–[29] are mostly designed to handle specific signal distortions present in digital images. The quantification of image quality for distorted images and filter-altered images may differ [11] [16]. More specifically, the former IQA focuses on the effect of geometric deformation on image structure and quantifies the perceptual disparity between a distorted image and its form of reference (i.e., either actual or proxy reference) [1]; the latter IQA focuses on the effect of filter altering on the visual properties of images and calculates the quality preference measure amongst visual stimuli. Obviously, building a computational model for the quality assessment of filter-altered images is rather challenging mainly due to the fact that the relationship between the filter strength or intensity and overall image quality tends to be non-monotonic and complex. Furthermore, the relationship might be different when alternative filters are applied. With the rapid growth of social media, it is urgently required that an effective method is in place to automatically evaluate the perceived quality of filter-altered images, optimising the quality of visual experiences of users. Therefore, there is a pressing need for both subjective and objective IQA studies.

In this paper, addressing the aforementioned challenges, we conduct an image quality subjective perception experiment for the edited image quality assessment based on a standard laboratory environment. Subsequently, we construct an IQA database (CUMAD) specifically for filter-altered images using this experimental setup. Additionally, we develop a dedicated NR-IQA model for filter-altered images, comprising a module for perceiving image styles and a novel module for integrating image style features with quality features. Experimental results validate the strong performance of our model.

The contributions of this work are summarised as follows:

- Firstly, we conduct a first-of-its-kind psychovisual exper-

Corresponding author: Jianxun Lou (jianxunlou@outlook.com)

Xinbo Wu, Jianxun Lou, Yingying Wu, Wanan Liu, Paul L. Rosin, Gualtiero Colombo, Stuart Allen, Roger Whitaker and Hantao Liu are with the School of Computer Science and Informatics, Cardiff University, CF24 4AG Cardiff, United Kingdom.

Jianxun Lou is also with the School of Computer Science, Northeast Electric Power University, Jilin, China.

Wanan Liu is also with the School of Management, Hangzhou Dianzi University, Hangzhou, 310018, China.

iment for the quality assessment of filter-altered images, involving 20 human subjects providing more than 7,000 image quality scores. This results in a new IQA database, named Cardiff University image Manipulation quality Assessment Database (CUMAD).

- Secondly, we perform statistical analyses on the subjective data of image quality assessment, revealing the perceptual behaviours of viewers on the filter-altered images.
- Finally, we construct a new computational IQA model, *Image Manipulation Quality Assessment (IMQA)*, to effectively assess the quality of filter-altered images. Experimental results demonstrate that the IMQA model significantly outperforms existing alternative IQA models.

II. Related Work

A. Subjective image quality assessment

In the literature, many subjective image quality experiments have been conducted to gain a fundamental understanding of how humans perceive and evaluate image quality. As a result, several image quality assessment (IQA) databases have been established to facilitate research, including the LIVE database [30] where participants were asked to assess the quality of images degraded by the JPEG, JPEG2000, white noise, Gaussian blur, and simulated Rayleigh fading channel distortions in a fully controlled lab environment. There are also larger-scale IQA databases collected using crowdsourcing, including TID2008 [12], KonIQ-10k [13], and LIVE In the Wild [16]. The TID2008 database comprises a total of 1700 images, encompassing various types of distortions such as noise, blur, compression, transmission errors, as well as brightness and contrast distortions. The KonIQ-10k database represents the in-the-wild database designed for ecological validity. Through crowdsourcing, KonIQ-10k collected image quality ratings from 1459 participants with 1.2 million assessments on 10073 images. The LIVE In the Wild Image Quality Challenge Database includes 1162 images captured from various mobile devices. It aims to study human perception of image distortions during capture, processing and storage. The LIVE In the Wild database also assesses the widely diverse authentic image distortions, whereas the TID2013 and KADID-10K databases evaluate the quality of artificially distorted images. However, the aforementioned IQA databases have their limitations. All these databases focus on distorted images (either artificially or authentically), and the image quality perception of filter-altered, high-quality images where signal distortions do not occur is largely unexplored. Also, it should be noted that while crowdsourcing methods can help generate larger-scale databases, the degree of perceptual relevance is often compromised in an uncontrolled data collection environment. For example, the fine-grained difference in image quality perception cannot be captured as the difference may be caused by numerous uncontrolled variables. Therefore, to gain a full understanding of the image quality perception of filter-altered images, we conduct subjective experiments in a fully-controlled lab environment to faithfully reflect viewers' perceptual outcomes.

B. Computational models

Building upon the subjective image quality assessment (IQA) and its databases, various computational IQA models have been developed for automated evaluation of image quality. IQA models are generally classified into full-reference (FR), reduced-reference (RR) and no-reference (NR) IQA depending on the extent to which the pristine reference is used. Also, IQA models can be classified into traditional IQA that reply on handcrafted features and deep learning-based IQA that use deep neural networks to automatically learn relevant features. Popular traditional FR-IQA models include SIM [11], MS-SSIM [31], and VIF [22]. There are deep learning-based FR-IQA models such as DR-IQA [24] and IQT [25]. With the growth of digital media technology, there has been a need to develop NR-IQA models. Traditional NR-IQA models are mainly categorised into those based on natural scene statistics (NSS) [15], [23], [32]–[34] and those based on feature engineering [35]–[38]. With the advancement and widespread adoption of deep learning, new techniques have been employed to enhance the performance of IQA models. For example, models such as WaDIQaM [39] and DB-CNN [27] demonstrate the effectiveness of convolutional neural networks in extracting image quality features. HyperIQA [28] utilises a ResNet network to categorise features into low-level and high-level features, transforming the latter to redirect the former. MetaIQA [40] employs meta-learning to train networks for individual distortion types, facilitating the learning of prior knowledge. MANIQA [29] proposes the extraction of image features through ViT and the generation of image quality scores through a multi-dimensional attention network. DEIQT [41] introduces a data-efficient image quality transformer capable of accurately assessing image quality with reduced data requirements. StairIQA [42] proposes a staircase structure to enhance IQA tasks by integrating features hierarchically and an iterative mixed database training strategy for training on multiple databases to improve generalisation. AGAIQA [43] introduces an adaptive graph attention module that optimises feature utilisation, along with a patch-wise hierarchical perceptual regression module for enhanced scoring accuracy. Existing IQA models are predominantly designed for assessing distorted images, and the applicability of these models for filter-altered images remains unknown hence will be investigated in this paper. As such, we aim to develop a dedicated IQA model for image manipulation quality assessment.

III. Psychovisual study

To understand viewers' image quality perception of the filter-altered images, we conduct a psychovisual study under a fully-controlled lab environment. The study results in a first-of-its-kind image quality database, the Cardiff University image Manipulation quality Assessment Database (CUMAD), containing high-quality natural images altered using popular image manipulation methods.

A. Stimuli

We systematically selected 60 high-quality, high-resolution (all resized to 1920 × 1080) source images from an open source image sharing website, Unsplash [44]. The dataset of source images contains a total of 10 categories of scenes including (1) Action (AC): images that show high activity, (2) Animal (AN): animal-themed images, (3) Food (FO): images of food and drink, (4) Indoor (IN): images captured from the indoor scenarios, (5) Night (NI): images captured at night, (6) Object (OB): images of various objects, (7) Outdoor Man-made (OM): images captured from outdoor scenarios with man-made objects, (8) Outdoor Natural (ON): images captured from outdoor scenarios with nature scenes, (9) Portrait (PO): close-up shots of human faces, (10) Social (SO): images with interactions between people. Fig.1 illustrates the source images from our study.

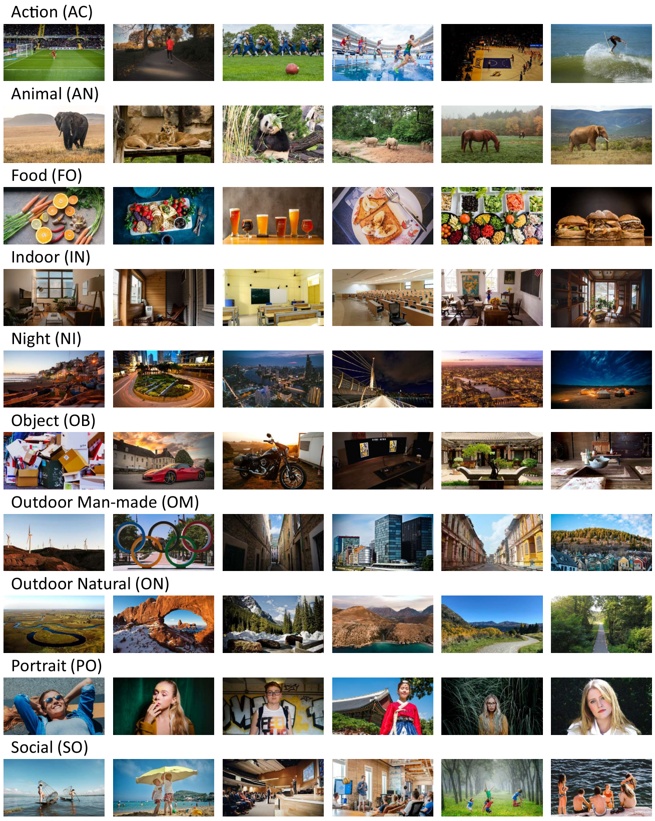


Fig. 1. Source images of the psychovisual study.

There are many image manipulation (IM) methods available on existing social media platforms and mobile applications. To gain a realistic study we have selected three popular methods for manipulating the source images. These IM methods can significantly transform the appearance of an image, enhancing its visual appeal, including Contrast Limited Adaptive Histogram Equalization (i.e., referred to as CLAHE), White Patch Retinex (i.e., referred to as PR) – a white balance adjustment method, and Image Toning (i.e., referred to as Toning). For each IM method, two editing versions were applied to each source image, reflecting two distinct levels of perceived overall quality. For the CLAHE and PR methods, we use CLAHE_1 and PR_1 to simulate subtle or moderate changes; and CLAHE_2 and PR_2 to simulate drastic or aggressive alterations. We used MATLAB software to manipulate the source images with the CLANE and PR methods. For the Toning method, we have selected two widely used methods of image toning, i.e., Toning_1 to simulate complementary colours and Toning_2 to simulate analogous colours. We used professional photo editing software (i.e., PaintShop Pro 2020), based on the principle of complementary colours and analogous colours, to edit the source images. As shown in Fig.2, for each source image, six different filter-altered images are generated by algorithmic and software processing. During the construction of the database, image quality experts conducted visual inspections and adjusted IM parameters to ensure that the six filter-altered images per source image were perceptually different from each other. As a result, a total of 360 stimuli were generated (excluding the source images) for the CUMAD database.

B. *Perception experiment*

We conducted a psychovisual experiment in the Visual Computing laboratory at Cardiff University. We set up a standard office environment for subjective image quality assessment as per the International Telecommunication Union (ITU) standards [45]. The laboratory provides a fully controlled experimental environment, ensuring consistent viewing conditions, involving characteristics such as low surface reflectance and constant ambient light. The visual stimuli were displayed on a 27-inch OLED monitor (native resolution is 2560×1440 pixels at 60 Hz). The viewing distance was set to be approximately 60cm. To ensure that the monitor displayed the correct colours in the lab environment, we calibrated the monitor using the SpyderX Elite monitor calibrator. We employed the standardised single-stimulus experimental protocol [45], [46], which allows participants to evaluate the quality of an image without a direct comparison to a reference image. Moreover, we adopted a within-subjects experiment design [47], in which all participants were each requested to view and score the entire set of test stimuli. It should be noted that although the within-subjects method can produce reliable IQA ratings, this is subject to carry-over effects due to fatigue and boredom for example [48]. To eliminate the carry-over effects in our experiment, the 360 test images were randomly divided into two partitions of 180 images each. Each subject was asked to complete two sessions (i.e., each session involves assessing 180 images) with a "break" period of 60 minutes between sessions. The test images were shuffled for each subject to ensure that the order of the stimuli presented to participants was randomised. This helps prevent order effects by reducing the likelihood of any systematic bias due to the presentation sequence. We recruited 20 participants in our experiment, consisting of 10 males and 10 females (between 19 to 50 years of age), all inexperienced with image quality assessment. The participants were not tested for vision defects, and we considered their verbal expression of the soundness of vision to be adequate. Each subject was provided with instructions on the purpose and general procedure of the experiment, and a training session (to familiarise subjects with the stimuli and scoring scale) before the start of the actual experiment. We used 12 images (i.e., 2 source images × 6 filter-altered images)

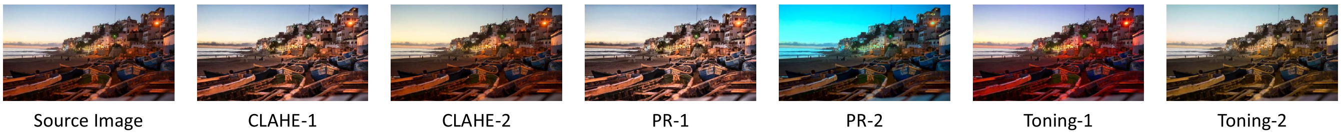


Fig. 2. Illustration of a source image and six different filter-altered images. ‘CLAHE-1’, ‘CLAHE-2’, ‘PR-1’, ‘PR-2’, ‘Toning-1’, and ‘Toning-2’ represent three popular image manipulation (IM) methods and each applied at two distinct levels of strength.

in the training session that were different from those used in the actual experiment. We did not limit the observation time for each stimulus; participants were allowed to observe the image until they felt they could express their opinion on the image quality scale. Informed consent was obtained from participants, and their privacy and confidentiality were rigorously protected throughout the research process.

C. Processing of subjective data

In scoring image quality, participants often use different segments of the scale to articulate their ratings. To account for the differences between individuals in using the scale, we standardise subjects’ scoring results to the same mean and standard deviation. In other words, we convert raw subjective scores to z-scores using the following formula:

$$z_{ij} = (r_{ij} - \mu_i) / \sigma_i, \tag{1}$$

where r_{ij} represents the raw score given by the i-th subject for the j-th test stimulus, μ_i is the mean of all scores for subject i, and σ_i is the standard deviation. Subsequently, the mean opinion score (MOS) for each stimulus is computed as the mean of the remaining z-scores across all subjects as follows:

$$MOS_j = F_{scale}\left(\frac{1}{s}\sum_{i=1}^{s} z_{ij}\right), \tag{2}$$

where s represents the number of remaining subjects for the j-th image. We also linearly transform the resulting MOS values to the range between 0 to 1. These data represent the Cardiff University image Manipulation quality Assessment Database (CUMAD), which consists of 60 diverse visual scenes and 360 filter-altered images with their perceived quality rated by 20 subjects.

D. Properties of CUMAD database

To analyse viewers’ image quality assessment (IQA) behaviour, we quantify the variation in scoring image quality between subjects. This involves calculating the Pearson linear correlation coefficient (PLCC) between each subject’s scoring results and the MOS values over the entire dataset. Fig.3 shows the PLCC values for individual subjects and the average PLCC value. The narrow 95% confidence interval of the correlation (i.e., [49]) suggests a high agreement between subjects in scoring the quality of test images.

We analyse the impact of scene category on the perceived image quality. Fig. 4 illustrates the category-wise MOS for 10 different scene categories contained in the CUMAD database. The same image filtering effects and parameters are applied to each scene category. However, their impact on the perceived image quality differs as shown in Fig. 4. For example, the image quality of the ON, OM and NI categories tends to be higher than for other categories, and the images of AC and PO categories are rated rather low in the CUMAD database. We perform hypothesis testing to further analyse the observed tendencies. An analysis of variance (ANOVA) is conducted by selecting image quality as the dependent variable and the scene category as the independent variable. The results of ANOVA show that the scene category has a statistically significant effect on the image quality (p-value<0.05 at 95% level).

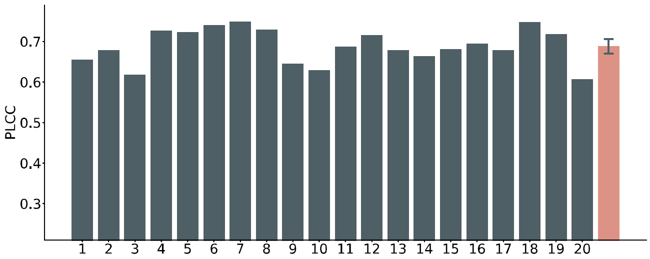


Fig. 3. Illustration of the correlation between mean opinion scores (MOS) and individual subjects’ scores on the CUMAD database. The right-most bar shows the mean correlation with a 95% confidence interval.

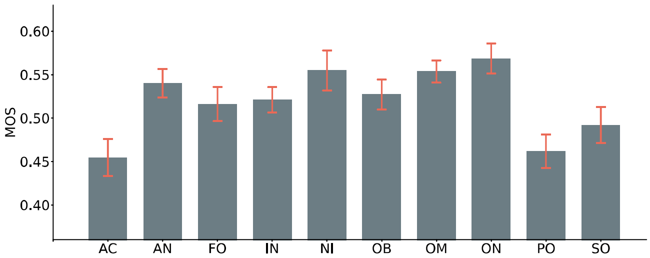


Fig. 4. The mean opinion score (MOS) of different scene categories contained in the the CUMAD database.

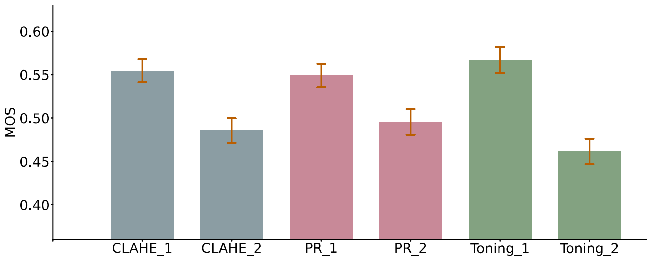


Fig. 5. The mean opinion score (MOS) of different image manipulation methods (type and strength level) contained in the the CUMAD database.

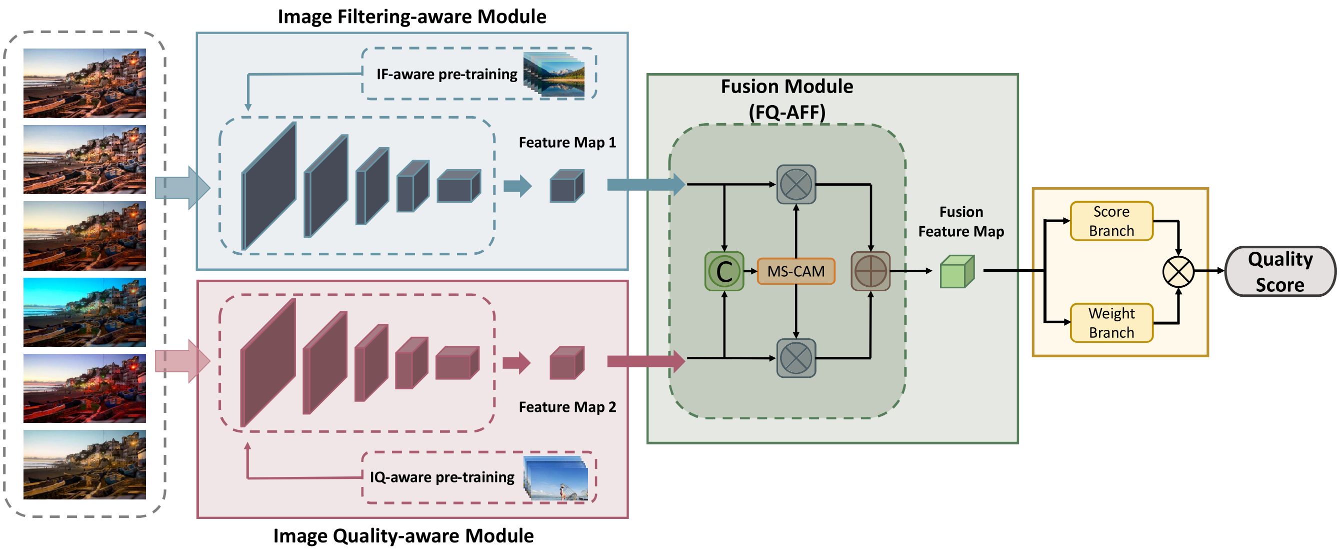


Fig. 6. Schematic overview of the proposed Image Manipulation Quality Assessment (IMQA) model. The input image is processed through both the image filtering-aware (IF-aware) encoder and the image quality-aware (IQ-aware) encoder. The resulting feature maps are then fused via a feature fusion module, and the fused feature map is fed into the patch-weighted quality predictor, yielding the final image quality assessment.

In the CUMAD database, two different editing strengths per image manipulation (IM) method are applied to each source stimulus. We analyse whether these two pre-defined strength levels actually produce two distinct levels of perceived quality in the subjective experiment. Fig.5 illustrates the MOS for each IM method with two different strength level. It can be seen that the MOS of CLAHE_1, PR_1 and Toning_2 are higher than that of CLAHE_2, PR_2 and Toning_1, respectively. This indicates that the strength of image filtering tends to cause the changes in the perception of image quality. To verify the observed tendencies on the impact of IM strength, an ANOVA is performed by selecting image quality as the dependent variable and the categorical IM strength level as the independent variable. The ANOVA results show that the IM strength has a statistically significant effect on the image quality (p-value<0.05 at 95% level).

IV. The Proposed Method

In contrast to the prevalence of filtered-altered images across various social media and mobile applications, there is a lack of dedicated computational models for automatic prediction of quality of filter-altered images. In this paper, we propose an Image Manipulation Quality Assessment (IMQA) model that aims to address the IQA problem beyond the perception of visual distortions and focus on the impact of image manipulation via filtering. The schematic overview of the proposed IMQA model is shown in Fig.6, which consists of four key components: (1) an image filtering-aware (IF-aware) encoder; (2) an image quality-aware (IQ-aware) encoder; (3) a feature fusion module, and (4) a patch-weighted quality predictor [29]. These components are detailed below.

A. *Image filtering-aware (IF-aware) and image quality-aware (IQ-aware) encoders*

The impact of image manipulation via filtering on perceived quality is multifaceted [50] and influenced by factors such as the type of filter applied, the strength of filtering effect, and visual scene context, as analysed for the ground truth of the CUMAD database in the above section. To conceptualise these plausible influencing factors in a deep learning-based IQA framework, we consider two types of features: the image filtering-aware (IF-aware) features that target the IQA representations of varying effects created by different types of filters; and image quality-aware (IQ-aware) features that form the IQA representations of basic image properties such as contrast and sharpness. Both IF-aware and IQ-aware features take into account the interplay between visual content and the specific features (i.e., filtering effects or quality attributes). For instance, contrast adjustments applied to landscape images often enhance the overall quality, while employing the same processing on portrait images may degrade the perceived quality, depending on the intensity [51]–[53].

To construct a deep learning-based model, we employ two specialised encoders, namely the image filtering-aware (IF-aware) encoder and the image quality-aware (IQ-aware) encoder, with each encoder being tailored to achieve a specific feature learning target. The IF-aware encoder is used for extracting deep features related to distinct filter types, while the IQ-aware encoder is used for extracting deep features related to image quality attributes. Both encoders have a similar structure and are constructed based on ConvNeXt [54] due to its proven performance in various computer vision tasks [54]. We remove the classification head from ConvNeXt to make it a feature extractor, and append an additional convolution layer at its end to reduce the number of channels of feature maps, resulting in lower computational complexity. Let $\mathcal{F}_{\text{filter}}$ and $\mathcal{F}_{\text{quality}}$ to denote the output feature maps from the IF-aware encoder and IQ-aware encoder, respectively. Given an input image $I \in \mathbb{R}^{3\times H\times W}$, the model produces feature maps $\mathcal{F}_{\text{filter}}, \mathcal{F}_{\text{quality}} \in \mathbb{R}^{768} \times \frac{H}{32} \times \frac{W}{32}$. Here, H and W denote the height and width of the input image, respectively. We design a dedicated training framework to effectively extract filter-related and quality-related features (i.e., $\mathcal{F}_{\text{filter}}$ and $\mathcal{F}_{\text{quality}}$) from the input images. The details of the training framework are described in section IV-D.

B. Feature fusion module

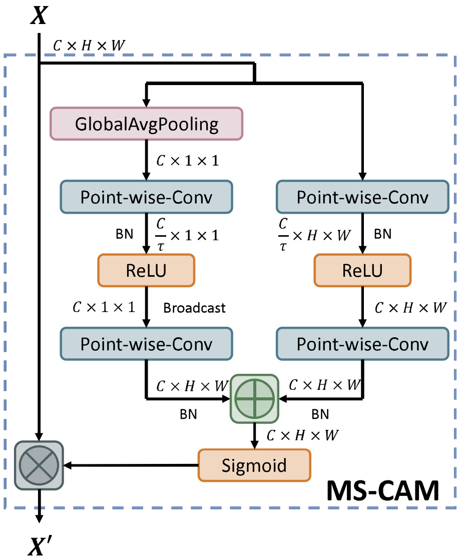


Fig. 7. Multi-Scale Channel Attention Module (MS-CAM) within the proposed Filter-Quality Attention Feature Fusion (FQ-AFF) module. The input feature map X is processed through both the global channel ($G(X)$) and the local channel ($L(X)$) to extract global and local features, respectively. These features are fused to produce attention weights for the original feature map X to enhance feature representations.

Since the perceived quality of filter-altered images is strongly influenced by both the filter type and conventional quality attributes, effective integration of the representations of the IF-aware and IQ-aware encoders is crucial for the overall model. We propose a Filter-Quality Attention Feature Fusion (FQ-AFF) module to fuse the $\mathcal{F}_{\text{filter}}$ and $\mathcal{F}_{\text{quality}}$ maps generated from the IF-Aware encoder and IQ-aware encoder. The structure detail of the FQ-AFF is illustrated in Fig.6. In FQ-AFF, the $\mathcal{F}_{\text{filter}}$ and $\mathcal{F}_{\text{quality}}$ maps are first concatenated, then fed into the Multi-Scale Channel Attention Module (MS-CAM) [55] to generate attention weights (X') for $\mathcal{F}_{\text{filter}}$ and $\mathcal{F}_{\text{quality}}$, respectively. Given that human visual system (HVS) process involves the perception of both global information (e.g., content, composition, white balance, and tone) and local information (e.g., texture, brightness, and sharpness), the MS-CAM is designed to simulate the HVS process, generating attention weights from both global and local perspectives. The diagram of the MS-CAM architecture is depicted in Fig.7, and its process can be expressed as follows:

$$X' = \sigma(G(\mathcal{F}_{\text{filter}} \uplus \mathcal{F}_{\text{quality}}) \oplus L(\mathcal{F}_{\text{filter}} \uplus \mathcal{F}_{\text{quality}})), \tag{3}$$

where the symbol $\oplus$ denotes the addition; $\uplus$ represents the concatenate operation. In this context, τ in Fig.7 represents the scaling factor for the channels. The MS-CAM takes a feature map X as input, which represents the concatenated features of $\mathcal{F}_{\text{filter}}$ and $\mathcal{F}_{\text{quality}}$. The two branches of the MS-CAM module represent the global channel context $G(X)$ and the local channel context $L(X)$, respectively. The outputs are fused to produce attention weights for the original input feature map X to enhance feature representation. Finally, the fused features (Z) can be represented as:

$$Z = X' \otimes \ \mathcal{F}_{\text{filter}} + X' \otimes \ \mathcal{F}_{\text{quality}}, \tag{4}$$

where $\otimes$ represents element-wise multiplication, also known as the Hadamard product.

C. Patch-weighted quality predictor

In [29], a dual-branch structure for patch-weighted quality predictor is proposed, demonstrating strong prediction performance. Given the feature F, the module generates weight (W) and score (S) projections. An effective approach taken in [29] is to treat the overall image quality as the aggregation of quality estimations of image patches. In this case, the final quality score for an image is the sum of the product of the score and weight for each patch. We implement this dual-branch structure in our model and fed it with the feature map Z obtained from the FQ-AFF module (IV-B). This process can be expressed as follows:

$$S_{out} = \sum_{i=1}^{n}(W_Z(i) \times S_Z(i)), \tag{5}$$

where S_{out} represents the final predicted image quality score. The feature map Z is divided into n patches, $W_Z(i)$ denotes the weight of the i-th patch in Z, and $S_Z(i)$ denotes the score of the i-th patch in Z.

D. Proposed datasets and pre-training strategy

Given that deep learning is a data-driven approach, employing appropriate datasets and pre-training strategies can significantly enhance the model's ability to achieve specific objectives [56]. To this end, we propose dedicated datasets and pre-training approaches tailored specifically to individual encoders, i.e., the IF-aware encoder and IQ-aware encoder. This enables an effective extraction of IF-Aware and IQ-Aware features for the assessment of filter-altered images.

To facilitate pre-training for the IF-aware encoder, we construct a large-scale Image Filtering-Aware Dataset (IFAD) to augment learning discriminative IF-aware features. Based on the six image manipulation (IM) methods as described in section III-A, 30,000 filter-altered images from a random collection of 5,000 images sourced from the Unsplash website [44] are generated. Fig. 8 (a) illustrates some sample images contained in the IFAD dataset. In pre-training, the IF-aware encoder is jointed with a classification head (i.e., a global average pooling with a fully connected layer), constituting a classification network that aims to classify different IM methods for the input images. The schematic overview of the process is illustrated in Fig. 8 (a). After pre-training on the

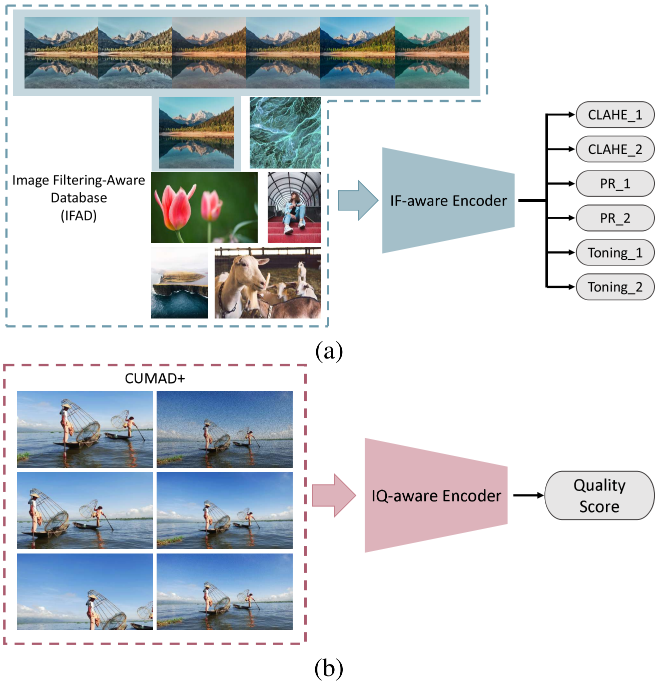


Fig. 8. Illustration of proposed dedicated datasets and pre-training strategy. (a) IF-aware encoder pre-training process. (b) IQ-aware encoder pre-training process.

IFAD dataset, the IF-aware encoder can effectively produce the image filtering related features.

For the pre-training of the IQ-aware encoder, we introduce the CUMAD+ dataset, tailored to enabling the encoder to extract features relevant to conventional image quality attributes. The CUMAD+ dataset comprises 9,000 stimuli and their corresponding quality scores. The stimuli consist of cropped and distorted variants of images from the CUMAD database, as some sample images illustrated in Fig. 8 (b). Cropping is a commonly employed technique in data augmentation [57]. To mitigate the risk of excessive cropping that may lead to the loss of semantic information and alterations of the aspect ratio, each image is randomly cropped to two-thirds and three-quarters of the original size while maintaining the original proportions [58]. This cropping process is repeated ten times for each specified cropping size to create sufficient diversity in the results. Meanwhile, to simulate varying levels of visual degradation the distorted variants are created by applying four common types of noise (Gaussian blur, motion blur, Gaussian noise, and salt-and-pepper noise) [57] to each image. The quality scores for the cropped images are assumed to be the same as those of their original counterparts contained in the CUMAD database, while the quality scores for the distorted images are derived from the average scores given by two state-of-the-art IQA models, i.e., MANIQA [29] and HyperIQA [28]. To guide the feature learning of the IQ-aware encoder, we add a regression head to its output layer for pre-training on the CUMAD+ database, as illustrated in Fig. 8 (b). Once the IF-aware encoder and IQ-aware encoder are pre-trained, we discard the auxiliary components, i.e., the classification head and regression head, and utilise the pre-trained encoders to constitute the architecture of the proposed IMQA model.

V. Experiment results

A. Implementation details

As mentioned in section IV-D, the IF-aware and IQ-aware encoders are first pre-trained on the dedicated IFAD dataset and CUMAD+ dataset, respectively. Subsequently, the pre-trained encoders are integrated to the final IMQA model as illustrated in Fig. 6 for fine-tuning towards the ultimate task of image manipulation quality assessment using the CUMAD database. To ensure comprehensive and fair reporting of result in the study, we employ k-fold cross-validation (k=10). More specifically, we divide the CUMAD database into ten non-overlapping subsets, each containing 36 images. In each iteration, one subset is held out as the test set, another as the validation set, and the remaining eight subsets are used as the training set. To eliminate unnecessary randomness, each test set corresponds to a fixed validation set. Additionally, to reduce biases in model learning, we ensure that each subset contains at least six different image manipulation (IM) methods and ten distinct scene categories to cover the full spectrum of IM variations. Our experiments were conducted on NVIDIA GeForce RTX 4090 equipped with PyTorch 1.13.1 and CUDA 11.7 for training and testing purposes. To reduce computational costs, all input images were resized to 480×270 pixels, maintaining the same aspect ratio as the original images (i.e., 16:9). The Adam optimiser [59] was employed to minimise the loss function. The learning rate was set to 5×10^{-6} and decayed by a factor of 0.9 after each epoch. The model was trained for 80 epochs with a batch size of 4.

B. Experiment results

To evaluate the performance of the predictive models, we utilise standards metrics, namely the Pearson Linear Correlation Coefficient (PLCC) [66], Spearman Rank Order Correlation Coefficient (SROCC) [67] and Root Mean Squared Error (RMSE) [68]. Both PLCC and SROCC metrics are within the range of [-1, 1], where values closer to -1 or 1 indicate higher correlation and better performance. Conversely, values that are closer to 0 suggest lower correlation and poorer performance. The RMSE metric ranges from 0 to positive infinity. A value of 0 indicates perfect alignment between the predicted results and the ground truth. In general, a lower RMSE value signifies better predictive performance of the model.

In this study, we compare the performance of our proposed IMQA model to 18 state-of-the-art (SOTA) NR-IQA [46] methods on the CUMAD database. It should be noted that only NR-IQA methods are relevant in our context as a reference for image manipulation quality assessment is nonexistent. These methods include traditional handcrafted IQA (HC-IQA) methods: BRISQUE [15], DipIQ [35], NIQE [23], and IL-NIQE [60]; deep learning-based IQA (DL-IQA) methods: P2P-BM [61], MetaIQA [40], WaDIQaM [39], DBCNN [27], UNIQUE [62], TRES [63], MANIQA [29], HyperIQA [28], DEIQT [41], StairIQA [42] and AGAIQA [43]; as well as deep learning-based image aesthetics assessment (DL-IAA)

TABLE I

Performance comparison of the proposed IMQA model to state-of-the-art (SOTA) models on the CUMAD benchmark database. HC-IQA: traditional handcrafted image quality assessment; DL-IQA: deep learning-based image quality assessment; and DL-IAA: deep learning-based image aesthetics assessment.

Models		Correlation IQA		
	Type	PLCC	SROCC	RMSE
BRISQUE [15]	HC-IQA	0.1307	0.0892	0.2747
DipIQ [35]	HC-IQA	0.1598	0.1758	0.2377
NIQE [23]	HC-IQA	0.2121	0.2204	0.3251
ILNIQE [60]	HC-IQA	0.2201	0.2291	0.3709
P2P-BM [61]	DL-IQA	0.1595	0.1432	0.2824
MetaIQA [40]	DL-IQA	0.3307	0.3199	0.2207
WaDIQaM [39]	DL-IQA	0.3612	0.2773	0.2811
DBCNN [27]	DL-IQA	0.3655	0.3824	0.2126
UNIQUE [62]	DL-IQA	0.5017	0.5001	0.1772
DEIQT [41]	DL-IQA	0.5132	0.5941	0.1233
MANIQA [29]	DL-IQA	0.5905	0.6003	0.1186
TRES [63]	DL-IQA	0.6035	0.5618	0.1181
StairIQA [42]	DL-IQA	0.6232	0.6332	0.1173
AGAIQA [43]	DL-IQA	0.6291	0.6711	0.1101
HyperIQA [28]	DL-IQA	0.6332	0.6094	0.1088
NIMA [26]	DL-IAA	0.5080	0.4620	0.1804
RAPID [64]	DL-IAA	0.5497	0.5356	0.1522
DMANet [65]	DL-IAA	0.6679	0.4647	0.1077
IMQA (ours)		**0.7253**	**0.6870**	**0.1003**

methods (note, IAA is relevant as the methods target the assessment of higher-quality images without conventional visual distortions): NIMA [26], RAPID [64], and DMANet [65]. These models were selected in our study as they represent the best-performing methods in popular IQA or IAA benchmarks (e.g., LIVE [22], TID2013 [69], AVA [70]). More importantly, we limited our selection to the models that have made their code transparently available for public use so we can perform a fair comparative study by implementing all models under the same experimental conditions.

Table I illustrates the performance comparison of our IMQA to traditional handcrafted IQA methods, deep learning-based IQA methods, and deep learning-based IAA methods on the CUMAD database. From Table I, it can be observed that our method outperforms existing methods by a large margin, in terms of both PLCC, SROCC and RMSE. The performance of traditional handcrafted IQA methods (e.g., BRISQUE, DipIQ, NIQE, IL-NIQE) is rather poor, indicating that these methods cannot be used to assess the perceived quality of filter-altered images. This is due to the fact that these methods are specifically designed for the assessment of distorted images, and hence cannot capture relevant IQA features for filter-altered images.

The performance of deep learning-based IQA methods tends to be better than that of the handcrafted IQA methods. For example, MANIQA, TRES, and HyperIQA demonstrate good performance, as they can fairly adapt to various image scenarios due to the powerful deep feature extractors built into these models. On the other hand, deep learning-based IAA methods exhibit even better performance, with DMANet achieving a PLCC score of 0.6679 (i.e., the best performance achieved by SOTA on the CUMAD database). These IAA methods were trained on a large-scale database for aesthetic visual analysis, AVA [70], implying that aesthetic-related features are useful for the quality assessment of filter-altered images.

The aforementioned results indicate the importance of including relevant and discriminative features for image manipulation quality assessment, and hence substantiate the rationale behind employing the IF-aware and IQ-aware encoders and dedicated pre-training strategies to extract representative features for the IMQA model.

To validate whether the observed performance gain of our proposed method is statistically significant, hypothesis testing is conducted. However, it should be noted that statistical testing is inherently complex, and the validity of conclusions depends upon the assumptions made about the underlying data. The choice of statistical methods can significantly influence the results, necessitating careful interpretation. To ensure robustness, we employ two well-established approaches for significance testing commonly used in the image quality literature. (1) Assuming that the two populations are normally distributed, we perform hypothesis testing (i.e., referred to as HT_P1401) using a pairwise comparison of the RMSE values (i.e., our IMQA model versus each SOTA model), following the procedure outlined in ITU-T Rec. P.1401 [71]. (2) Hypothesis testing (i.e., referred to as HT_Res) is performed based on the residuals between each SOTA model and our proposed IMQA, as described in [72]. We first assess the normality assumption of the residuals (i.e., $|MOS - NR\_IQA|$ and $|MOS - IMQA|$). When the two sets of residuals follow a normal distribution, an independent samples t-test is conducted; otherwise, the Mann-Whitney U test is performed. The results from both the HT_P1401 and HT_Res tests are shown in Table I, demonstrating that the performance improvement from our proposed IMQA model is statistically significant as compared to any of the SOTA models.

C. Ablation study

We conduct a comprehensive ablation study to quantify the contribution of individual components of the proposed IMQA model. This includes (1) the IF-aware encoder, (2) the IQ-aware encoder, and (3) the FQ-AFF feature fusion module.

To this end, we created five model variants: Baseline – IMQA that only contains a single encoder (i.e., ConvNeXt without dedicated pre-training); Variants *A* – IMQA with IQ-aware encoder only; Variants *B* – IMQA with IF-aware encoder only; Variants *C* – IMQA with both IQ-aware and IF-aware encoders but using simple feature fusion (i.e., concatenation); and Variants *D* – the final proposed IMQA model.

Table III illustrates the results of the ablation study. By comparing Variants *A* and *B* to the Baseline model, the benefit of including either the IQ-aware encoder or the IF-aware encoder to the IMQA model is clear. This highlights the importance of both conventional image quality features and image filtering related features in assessing the quality of filter-altered images. By comparing Variant *C* with Variants *A* and *B*, it is evident that the combined use of the IF-aware and IQ-aware encoders enhances the performance of the overall model. By employing the FQ-AFF rather than a simple feature

TABLE II

STATISTICAL SIGNIFICANCE TESTING, BASED ON TWO WELL-ESTABLISHED APPROACHES. (1) HYPOTHESIS TESTING (I.E., REFERRED TO AS HT_P1401) USING A PAIRWISE COMPARISON OF THE RMSE VALUES BETWEEN EACH SOTA MODEL AND OUR IMQA MODEL, AS OUTLINED IN ITU-T REC. P.1401 [71]. (2) HYPOTHESIS TESTING (I.E., REFERRED TO AS HT_RES) BASED ON THE RESIDUALS BETWEEN EACH SOTA MODEL AND OUR PROPOSED IMQA, AS DESCRIBED IN [72]. THE CODEWORD REPRESENTS THE TEST RESULTS FOR HT_P1401 AND HT_RES: "1" MEANS THAT THE DIFFERENCE IN PERFORMANCE BETWEEN A SOTA MODEL AND OUR IMQA MODEL IS STATISTICALLY SIGNIFICANT; "0" MEANS THAT THE DIFFERENCE IS NOT SIGNIFICANT.

	BRISQUE	DipIQ	NIQE	ILNIQE	P2P-BM	MetaIQA	WaDIQaM	DBCNN	UNIQUE
IMQA (ours)	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-1
	DEIQT	MANIQA	TRES	StairIQA	AGAIQA	HyperIQA	NIMA	RAPID	DMANet
IMQA (ours)	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-1	1-1

TABLE III

ABLATION STUDY TO QUANTIFY THE CONTRIBUTION OF INDIVIDUAL COMPONENTS OF THE PROPOSED IMQA MODEL. (**BOLD** FONT INDICATES THE BEST PERFORMANCE. ⊎ REPRESENTS SIMPLE FEATURE FUSION, I.E., CONCATENATION)

Variant	IQ-Aware Encoder	IF-Aware Encoder	Fusion Module	PLCC	SROCC
Baseline	–	–	–	0.4991	0.5052
A	✓	–	–	0.5507	0.5387
B	–	✓	–	0.6075	0.5825
C	✓	✓	⊎	0.6495	0.6440
D	✓	✓	FQ-AFF	**0.7253**	**0.6870**

fusion method, Variant *D* outperforms Variant *C* and achieves best performance in predicting image manipulation quality.

TABLE IV

ABLATION STUDY TO QUANTIFY THE CONTRIBUTION OF DEDICATED PRE-TRAINING TO THE PROPOSED IMQA MODEL. (**BOLD** FONT INDICATES THE BEST PERFORMANCE.)

Variant	IF-Aware Pre-training	IQ-Aware Pre-training	PLCC	SROCC
PT-Baseline	–	–	0.5338	0.5327
PT-A	✓	–	0.6671	0.6566
PT-B	–	✓	0.5901	0.5920
PT-C	✓	✓	**0.7253**	**0.6870**

In addition, we perform an ablation study to evaluate the contribution of dedicated pre-training to the proposed IMQA model. To this end, we created four model variants: PT-Baseline – IMQA without dedicated pre-training on encoders; Variants *PT-A* – IMQA with pre-training on the IF-aware encoder only; Variants *PT-B* – IMQA with pre-training on the IQ-aware encoder only; and Variants *PT-C* – IMQA with pre-training on both the IQ-aware and IF-aware encoders (i.e., the final proposed IMQA model).

As can be seen from the ablation study results in Table IV, employing dedicated pre-training on either the IQ-aware encoder or IF-aware encoder individually leads to a performance gain compared to the baseline model without encoder pre-training. Furthermore, it is clear that pre-training both encoders of the IMQA model results in best performance. This demonstrates the effectiveness of our proposed pre-training strategies for predicting the perceived quality of filter-altered images.

VI. DISCUSSION

Evaluating model complexity is crucial for selecting appropriate models for various real-world applications. Given that real-time responsiveness and model size are critical factors in practical settings, we measure the inference time and the number of parameters for the top-performing models to reflect their complexity. As per Table I, these top-performing models achieve PLCC and SROCC values above 0.5, and RMSE values below 0.2 on the CUMAD benchmark database. Table V illustrates the results of the complexity analysis. While our proposed model exhibits greater complexity compared to its competitors, its predictive performance, as detailed in section V-B, significantly surpasses that of the other models. In practical applications, maintaining an optimal balance between a model's complexity and performance is often necessary. This highlights the importance of evaluating the trade-offs based on the specific requirements of each use case.

TABLE V

COMPLEXITY ANALYSIS OF SOTA MODELS AND OUR PROPOSED IMQA MODEL ON THE CUMAD BENCHMARK DATABASE.

Model	Inference Time(per image)	Parameter Count
UNIQUE [62]	0.10s	22.3M
DEIQT [41]	0.05s	40.5M
MANIQA [29]	0.10s	135.6M
TRES [63]	0.06s	152.5M
StairIQA [42]	0.05s	122.5M
AGAIQA [43]	0.11s	148.8M
HyperIQA [28]	0.05s	27.4M
NIMA [26]	0.15s	134.3M
RAPID [64]	0.02s	2.1M
DMANet [65]	0.12s	130.3M
IMQA (ours)	0.17s	397.8M

VII. CONCLUSION

In this paper, we have conducted a psychovisual study to reveal how humans perceive the quality of filter-altered images. This study results in a first-of-its-kind benchmark database, named CUMAD, for image manipulation quality assessment. We have also developed a deep learning-based model, named IMQA, for automated quality assessment of filter-altered images. The IMQA model integrates the image filtering related features and image quality attributes related features to learn representations for filter-altered images. Experimental results demonstrate that the proposed IMQA model significantly outperforms state-of-the-art models in predicting the quality of filter-altered images.

REFERENCES

[1] N. Burningham, Z. Pizlo, and J. P. Allebach, "Image quality metrics," *Encycl. Imaging Sci. Technol.*, vol. 1, pp. 598–616, 2002.

[2] G. Yue, D. Cheng, T. Zhou, J. Hou, W. Liu, L. Xu, T. Wang, and J. Cheng, "Perceptual quality assessment of enhanced colonoscopy images: A benchmark dataset and an objective method," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 10, pp. 5549–5561, 2023.

[3] H. Qin and M. A. El-Yacoubi, "Deep representation for finger-vein image-quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1677–1693, 2018.

[4] J. Wen, F. Qin, J. Du, M. Fang, X. Wei, C. L. P. Chen, and P. Li, "Msg-Fusion: Medical semantic guided two-branch network for multimodal brain image fusion," *IEEE Trans. Multimed.*, pp. 1–14, 2023.

[5] C. Li, Z. Zhang, H. Wu, W. Sun, X. Min, X. Liu, G. Zhai, and W. Lin, "AGIQA-3K: An open database for ai-generated image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2023.

[6] Y. Yang, T. Xiang, S. Guo, X. Lv, H. Liu, and X. Liao, "EHNQ: Subjective and objective quality evaluation of enhanced night-time images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 4645–4659, 2023.

[7] A. Mikhailiuk, M. Pérez-Ortiz, D. Yue, W. Suen, and R. K. Mantiuk, "Consolidated dataset and metrics for high-dynamic-range image quality," *IEEE Trans. Multimed.*, vol. 24, pp. 2125–2138, 2022.

[8] A. Dziembowski, D. Mieloch, J. Stankowski, and A. Grzelka, "IV-PSNR—the objective quality metric for immersive video applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7575–7591, 2022.

[9] G. Wang, Z. Wang, K. Gu, K. Jiang, and Z. He, "Reference-free DIBR-synthesized video quality metric in spatial and temporal domains," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1119–1132, 2022.

[10] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synth. Lect. Image Video Multimed. Process.*, vol. 2, no. 1, pp. 11–15, 2006.

[11] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[12] 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," *Adv. Mod. Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.

[13] V. Hosu, H. Lin, T. Sziranyi, and D.Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 4041–4056, 2020.

[14] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imaging*, vol. 19, no. 1, pp. 011 006–011 006, 2010.

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

[16] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, 2015.

[17] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process. Image Commun.*, vol. 30, pp. 57–77, 2015.

[18] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A large-scale artificially distorted iqa database," in *2019 11th Int. Conf. Quality Multimed. Experience (QoMEX)*. IEEE, 2019, pp. 1–3.

[19] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[20] P. Teo and D. Heeger, "Perceptual image distortion," in *Proc. 1st Int. Conf. Image Process.*, vol. 2, 1994, pp. 982–986 vol.2.

[21] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imaging*, vol. 11, no. 2, pp. 206–223, 2002.

[22] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, 2006.

[23] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.

[24] A. Ghildyal and F. Liu, "Shift-tolerant perceptual similarity metric," in *Eur. Conf. Comput. Vis.* Springer, 2022, pp. 91–107.

[25] G. Yin, W. Wang, Z. Yuan, C. Han, W. Ji, S. Sun, and C. Wang, "Content-variant reference image quality assessment via knowledge distillation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 3, 2022, pp. 3134–3142.

[26] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, 2018.

[27] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, 2018.

[28] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3667–3676.

[29] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, and Y. Yang, "MANIQA: Multi-dimension attention network for no-reference image quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1191–1200.

[30] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, 2006.

[31] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *37th Asilomar Conf. Signals, Syst. Comput., 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[32] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, 2013.

[33] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *J. Vis.*, vol. 17, no. 1, pp. 32–32, 2017.

[34] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, 2010.

[35] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, 2017.

[36] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, 2012.

[37] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, 2012, pp. 1098–1105.

[38] P. Ye, J. Kumar, and D. Doermann, "Beyond human opinion scores: Blind image quality assessment based on synthetic scores," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4241–4248.

[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 Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, 2017.

[40] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "MetaIQA: Deep meta-learning for no-reference image quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14 143–14 152.

[41] G. Qin, R. Hu, Y. Liu, X. Zheng, H. Liu, X. Li, and Y. Zhang, "Data-efficient image quality assessment with attention-panel decoder," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 2, 2023, pp. 2091–2100.

[42] W. Sun, X. Min, D. Tu, S. Ma, and G. Zhai, "Blind quality assessment for in-the-wild images via hierarchical feature fusion and iterative mixed database training," *IEEE J. Sel. Top. Signal Process.*, vol. 17, no. 6, pp. 1178–1192, 2023.

[43] H. Wang, J. Liu, H. Tan, J. Lou, X. Liu, W. Zhou, and H. Liu, "Blind image quality assessment via adaptive graph attention," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2024.

[44] "Unsplash: Beautiful free images & pictures." [Online]. Available: https://unsplash.com/

[45] R. I.-R. BT, "Methodology for the subjective assessment of the quality of television pictures," *Int. Telecommun. Union*, 2002.

[46] T. Installations and L. Line, "Subjective video quality assessment methods for multimedia applications," *Networks*, vol. 910, no. 37, p. 5, 1999.

[47] G. Keren, "Between-or within-subjects design: A methodological dilemma," *A handbook for data analysis in the behaviorial sciences*, vol. 1, pp. 257–272, 2014.

[48] L. Lévêque, J. Yang, X. Yang, P. Guo, K. Dasalla, L. Li, Y. Wu, and H. Liu, "CUID: A new study of perceived image quality and its subjective assessment," in *2020 IEEE Int. Conf. Image Process. (ICIP)*, 2020, pp. 116–120.

[49] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011.

[50] T. Bennett, M. Savage, E. B. Silva, A. Warde, M. Gayo-Cal, and D. Wright, *Culture, class, distinction.* Routledge, 2009.

[51] Y. Luo and X. Tang, "Photo and video quality evaluation: Focusing on the subject," in *Comput. Vis. – ECCV 2008: Proc. 10th Eur. Conf. Comput. Vis., Marseille, Fr., Oct. 12-18, 2008, Part III.* Springer, 2008, pp. 386–399.

[52] R. D. Zakia and D. Page, *Photographic composition: A visual guide*. Routledge, 2012.

[53] S. Zeki, "Clive bell's "significant form" and the neurobiology of aesthetics," *Front. Hum. Neurosci.*, vol. 7, p. 730, 2013.

[54] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[55] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 3560–3569.

[56] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[57] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[58] R. Takahashi, T. Matsubara, and K. Uehara, "Data augmentation using random image cropping and patching for deep cnns," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 9, pp. 2917–2931, 2019.

[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.

[60] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, 2015.

[61] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, and A. Bovik, "From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3575–3585.

[62] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Trans. Image Process.*, vol. 30, pp. 3474–3486, Mar. 2021.

[63] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2022, pp. 1220–1230.

[64] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "RAPID: Rating pictorial aesthetics using deep learning," in *Proc. 22nd ACM Int. Conf. Multimed.*, 2014, pp. 457–466.

[65] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 990–998.

[66] K. Pearson, "Note on regression and inheritance in the case of two parents," *Proc. R. Soc. Lond.*, vol. 58, no. 347-352, pp. 240–242, 1895.

[67] C. Spearman, "The proof and measurement of association between two things," *Am. J. Psychol.*, vol. 100, no. 3/4, pp. 441–471, 1987.

[68] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

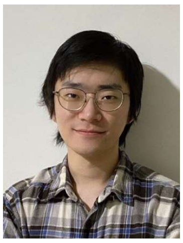
[69] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Image database tid2013: Peculiarities, results and perspectives," *Signal Process. Image Commun.*, vol. 30, pp. 57–77, 2015.

[70] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *2012 IEEE Conf. Comput. Vis. Pattern Recognit.* IEEE, 2012, pp. 2408–2415.

[71] "Statistical analysis, evaluation and reporting guidelines of quality measurements," International Telecommunication Union, Telecommunication standardization sector, Geneva, Tech. Rep. ITU-T P.1401, 2020.

[72] J. Antkowiak, T. J. Baina, F. V. Baroncini, N. Chateau, F. FranceTelecom, A. C. F. Pessoa, F. S. Colonnese, I. L. Contin, J. Caviedes, and F. Philips, "Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000," *Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000*, 2000.

Xinbo Wu received his M.S. and Ph.D. degrees from the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K., in 2020 and 2024. He has been the visiting scholar of Konstanz University, Konstanz, Germany. His research interests include visual quality assessment, visual perception and attention, and human-computer interaction.



Jianxun Lou received his B.Eng. degree from Central South University, Changsha, China, in 2018, and his M.S. and Ph.D. degrees from the School of Computer Science and Informatics at Cardiff University, Cardiff, U.K., in 2020 and 2024, respectively. His research interests include visual perception modelling and visual quality assessment.

Yingying Wu received the M.Sc. degree in data science and analytics from Cardiff University, U.K., in 2020, where she is currently pursuing the Ph.D. degree with the School of Computer Science and Informatics. Her research interests include image data analysis, human visual perception, and machine learning.

Wanan Liu received the Ph.D. degree from the Glorious Sun School of Business and Management, Donghua University, Shanghai, China, in 2023. He joined the School of Management at Hangzhou Dianzi University, Hangzhou, Zhejiang, China, after completing the doctoral studies. His principle research interests include business intelligence, data mining, and deep learning theories and its applications.

Paul L. Rosin is a Professor at the School of Computer Science & Informatics, Cardiff University. Previous posts include Brunel University, Joint Research Centre, Italy and Curtin University of Technology, Australia. His research interests include computer vision, remote sensing, mesh processing, non-photorealistic rendering, performance evaluation, shape analysis, facial analysis, and cultural heritage.

Gualtiero B. Colombo I am a Lecturer at Cardiff University where I have been previously working as a Researcher and Research Software Engineer. My research primarily concerns social computing, intelligence and evolution, and agent-based modelling by exploiting parallel processing at scale. From my PhD years in evolutionary optimisation I have built up extensive expertise for cross-disciplinary research projects spanning computer science, social sciences, and psychology. I supported model development and complex code for a range of projects including agent-based modelling and artificial intelligence deployed on high performance computing resources. My recent research interests focus on interpreting psychological and evolutionary concepts alongside the relation between AI, Innovation, and Creativity.

Professor Stuart Allen obtained an undergraduate degree in Mathematics from Nottingham University, and a PhD in Graph Theory from the University of Reading. Following his PhD, he moved to Cardiff in 1996 to begin a research project which developed new computational techniques for automated frequency assignment for military and commercial applications. His research broadened to address network design problems arising in 3G/4G cellular, emergency services, television and rural broadband, funded by the EPSRC, EU, industry and OFCOM. His current research interests cover interdisciplinary collaborations exploring the opportunities and challenges of emerging technologies, particularly links between psychology and smartphone use. He is currently a board member of Airbus Endeavr Wales, a joint initiative between Airbus and the Welsh Government to support innovation in Wales. Since 2015, Prof Allen has been the Head of Computer Science and Informatics at Cardiff University.

Roger Whitaker is currently a Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. His research interests include the intersection of machine and human intelligence, including human behavior. He is the Area Editor of Online Social Networks and Media (Elsevier) and an Associate Editor for Social Network Analysis and Mining (Springer).

Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands in 2011. He is currently a Professor at the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K.