Domain Fingerprints for No-reference Image Quality Assessment

Weihao Xia, Yujiu Yang, *Member, IEEE*, Jing-Hao Xue, Jing Xiao, *Member, IEEE*

Abstract—Human fingerprints are detailed and nearly unique markers of human identity. Such a unique and stable fingerprint is also left on each acquired image. It can reveal how an image was degraded during the image acquisition procedure and thus is closely related to the quality of an image. In this work, we propose a new no-reference image quality assessment (NR-IQA) approach called domain-aware IQA (DA-IQA), which for the first time introduces the concept of domain fingerprint to the NR-IQA field. The domain fingerprint of an image is learned from image collections of different degradations and then used as the unique characteristics to identify the degradation sources and assess the quality of the image. To this end, we design a new domain-aware architecture, which enables simultaneous determination of both the distortion sources and the quality of an image. With the distortion in an image better characterized, the image quality can be more accurately assessed, as verified by extensive experiments, which show that the proposed DA-IQA performs better than almost all the compared state-of-the-art NR-IQA methods.

Index Terms—No-reference image quality assessment, domain fingerprints, generative adversarial network

I. INTRODUCTION

NO-reference image quality assessment (NR-IQA) [1–4] is a fundamental yet challenging task that automatically assesses the perceptual quality of a degraded image without the corresponding reference for comparison. It serves as a key component in low-level computer vision, since in many applications it is difficult or even impossible to acquire the non-distorted image as reference to evaluate the quality of a distorted image. The ill-posed nature of NR-IQA is particularly pronounced for the absence of the prior knowledge about distortion form.

Numerous efforts [5–12] have been made to extract powerful features to represent images and image degradations. Traditional hand-crafted feature based methods usually leverage natural scene statistics (NSS) [2, 10, 11] and learning-based metrics [7–9]. For example, Saad *et al.* [11] leverage the statistical features of discrete cosine transform (DCT) for blind image quality assessment. Mittal *et al.* [2] propose to extract NSS features in the spatial domain to estimate the

W. Xia, Y. Yang are with Shenzhen International Graduate School, Tsinghua University, China. (Corresponding author: Yujiu Yang)

J.-H. Xue is with the Department of Statistical Science, University College London, UK.

J. Xiao is with Ping An Technology Co., Ltd., Shenzhen, China.

image quality. Ye *et al.* [7, 8] propose codebook representation approaches to predict subjective image quality scores by support machine regression (SVR). These hand-crafted features, however, lack flexibility and diversity for representing complex diverse degradations.

In recent years, deep learning methods achieve promising results in NR-IQA. Due to the extremely limited training datasets, many methods use various data augmentation or multi-task strategies to better exploit the power of Deep Neural Networks (DNNs). Some methods [13–15] focus on simulating the behavior of Human Visual System (HVS) by Generative Adversarial Networks (GANs). Specifically, Lin *et al.* [13] propose a discrepancy-guided quality regression network to encode the difference between distorted image and hallucinated reference to make precise prediction. Ren *et al.* [14] propose the restorative adversarial net, *i.e.*, the restorator reconstructs the non-distorted patches while the discriminator tries to distinguish the restored patches from the pristine distortion-free ones, and the evaluator takes the distorted patch and the restored patch as inputs and predicts a perceptually quantified score. However, these methods treat images of different distortions the same and the discriminative representations of distortions are under-explored.

Different types of distortion change distortion-free images into different distorted versions, leading to significantly different visual perception. As shown in Figure 1, the image quality depends on the distortion type and level. Some previous NR-IQA methods [16, 17] try to evaluate the image quality by considering the information of the distortion, but most of them simply use the distortion information by adding a classification to identify the distortion.

To address the above drawbacks, we propose the domain-aware no-reference image quality assessment (DA-IQA), which exploits the domain fingerprints for image quality assessment. A domain here is defined as the images with the same type and similar levels of distortion. Similar to human fingerprints, a unique and stable fingerprint is also left on each acquired image. It can reveal how an image was degraded during the image acquisition procedure and thus is closely related to the quality of an image. To get domain fingerprints, we design a novel domain-aware architecture to get disentangled representations for different domains. These representations are used as detailed and unique markers to better express particular degradation information, as shown in Figure 2. More specifically, a degraded image can be decomposed as the image content together with the degradation. We disentangle the image content and degradation features from degraded images to more accurately encode degradation information into the image

fastfading (44.62) Gaussian blur (46.89) JP2K (45.83)

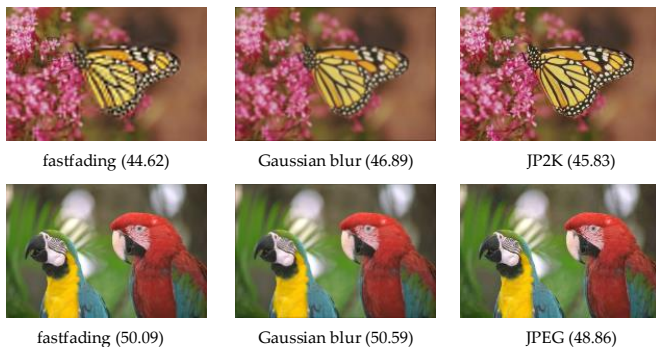fastfading (50.09) Gaussian blur (50.59) JPEG (48.86)

Fig. 1. **The two aspects of image quality.** The quality of an image is strongly linked to its distortion type. The degradation arises from the specific distortion, and the image quality depends on the distortion type and level. Higher score denotes worse quality (DMOS, range [0,100]).

quality assessment framework. The disentangled representations for different domains would be discriminated significantly well in high dimensional feature space. Furthermore, instead of adding a classification layer as in [16, 17], benefiting from the design of domain-aware network architecture, DA-IQA can also identify the distortion type of an image simultaneously.

Our contributions are summarized as follows:

- We propose the domain-aware no-reference image quality assessment (DA-IQA), which for the first time introduces the discriminative disentangled representations for different types of distortions to image quality assessment.
- Our method can also identify the distortion type of an image, and use the distortion type and quality score to characterize the image quality.
- Our method achieves superior performance on popular IQA datasets to state-of-the-art methods.

II. RELATED WORK

A. No-Reference Image Quality Assessment

The existing studies on NR-IQA can be broadly classified into two categories: designing hand-crafted features [1, 2, 10, 11] and learning discriminant visual features automatically [7–9]. The first category of methods typically use a two-stage framework, which performs the distortion identification and the distortion-specific quality estimation accordingly. However, Mittal *et al.* [2] have shown that such two-stage methods are not superior to the distortion-blind approaches. The second category of work attempts to learn discriminant visual features automatically without using hand-crafted features. Ye *et al.* [7, 8] construct a small yet accurate codebook to look up the proper features. Kang *et al.* [5, 16] and Bosse *et al.* [6] adopt deep neural network to extract features from the raw input and perform regression to estimate perceptual scores. The above NR-IQA methods can be summarized as feature extraction and regression based only on distorted images. However, according to the free-energy theory [18], HVS tends to restore the distorted image before quality assessment. Despite building NR-IQA models based on the free-energy theory, [3, 19] restore the distorted image with a linear autoregressive model, which is not capable of producing a satisfactory result

when the input suffers from high-level distortion and therefore may not be consistent with HVS. Lin *et al.* [13] and Ren *et al.* [14] simulate the behavior of HVS by using generative adversarial networks (GANs) to generate the corresponding restored counterparts as reference. Ren *et al.* [14] propose the restorative adversarial net and Lin *et al.* [13] propose a discrepancy-guided quality regression network to encode the difference between distorted image and hallucinated reference to make precise prediction. However, their methods do not exhibit the capability of disentangling and characterising discriminative latent representations for different degradations, which is one of our key contributions.

B. Representation Disentanglement

Many recent works on disentangled representation aim to learn an interpretable and transferable representation. For example, Denton *et al.* [20] separate time-independent and time-varying components for long-term video prediction. Some studies [21–23] focus on disentanglement of content and style to achieve multi-domain image translation. It is difficult to define content and style explicitly, and different studies adopt different definitions for their specific tasks. Liu *et al.* [24] propose a unified model that learns disentangled representation for describing and manipulating data across multiple domains. For image restoration, Lu *et al.* [25] disentangle the content and blur features from blurred images. Different from [25], we disentangle the content and discriminative representations of multiple degradations. We use the content features to restore images and use the discriminative representations for NR-IQA.

C. Image Fingerprint

Digital fingerprint is a signature that could be used to identify, track, monitor and monetize images by converting their content into compact digital asset or impression. Prior digital fingerprint techniques focus on detecting hand-crafted features for device fingerprints [26]. Recently, [27] introduces this concept to the image forensic field and show their application to the GAN source identification. Based on that, [28] replaces the hand-crafted fingerprint formulation with a learning-based one, and classify an image as real or GAN-generated by learning GAN fingerprints of different GAN models.

D. Domain-Aware Applications

The word *Domain-aware* mostly occurs in the NLP applications and its meaning varies from case to case. For example, [29] propose a domain-aware dialog system, which aims to maintain a fluent and natural conversation within the domain as well as during switching of domains. The domain in this case refers to the topic or theme of the conversation. Slightly different, [30] refer domain to a specific field such as retail, travel and entertainment, and introduce the task of multimodal domain-aware conversations.

A *domain* in our work refers to a collection of images with certain degradation. The process of image restoration can be formulated as the translation from the degraded domain to the pristine domain. Thus our work is mostly related to
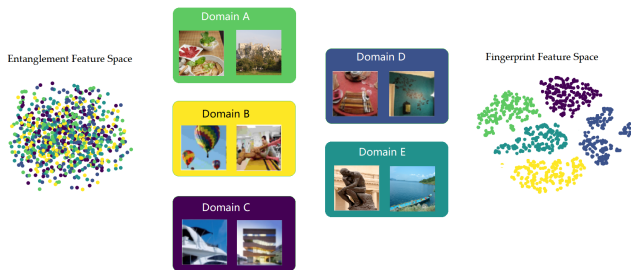
Fig. 2. **Illustration of domain fingerprints.** We use t-SNE [32] for visual comparison between the entangled features (left) and our fingerprint features (right), for degradation representations.

multi-domain image translation. Specifically, [31] recently proposes a unified model to achieve multi-domain image-to-image translation. [24] proposes a model that is able to perform continuous cross-domain image translation and exhibits ability to learn and disentangle desirable latent representations.

In this work, we demonstrate the existence and uniqueness of domain fingerprints, that signify disentangled discriminative representations of different degradations, for image quality assessment tasks.

III. OUR APPROACH

A. Problem Formulation

Given a distorted image I_d, our goal is to learn a mapping $f : I_d \rightarrow s$, in which $s \in \mathbb{R}^+$ denoted the predicted quality score of I_d and should be consistent with the result of Mean Opinion Score (MOS). To simulate the Human Visual System (HVS), we first restore degraded contents as reference. Assume in a dataset, there are n domains $\{\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_n\}$ of images. Each domain represents a collection of images with certain degradation or distortion. For each image I_d in a domain $\mathcal{D}$, it can be represented as the combination of pristine image I_{gt} and a certain distortion d, i.e., $I_d = I_{gt} \otimes d$. Our goal is to disentangle the representation d of a distorted image I_d to get the restored image I_r consistent with I_{gt}, and use this domain-aware distortion pattern to obtain image quality score s under the supervision of ground truth human visual quality score s_{gt}.

B. Overview of the Proposed Approach

The framework of our proposed DA-IQA is demonstrated in Figure 3. As shown, given several collections of images with different types and levels of degradations, referred to as n domains, an image I_d is randomly selected from one domain, the generator G tries to produce an image indistinguishable with the real pristine image to the image discriminator D. At the same time, the domain discriminator D_c tries to recognize the domain label c_i of representation and also reacts on the generator to further disentangle discriminative features for different degradations. After the convergence of restoration, we further train a regression network by using the high-level domain-aware distorted representation (domain fingerprint) and the obtained restored images I_r to get the

desired quality score s. The aforementioned processes are implemented by two corresponding modules, the domain-aware image restoration network and the hallucination-guided quality regression network, respectively, as shown in Figure 3.

C. Domain-Aware Image Restoration Network

The overview of proposed domain-aware image restoration network (DA-Restore) is shown in Figure 4. This DA-Restore module takes the distorted image as input and aims to produce the corresponding restoration. Meanwhile, the model tries to disentangle the distorted representation of specific degradation from the content. The restored image could act as a hallucinated reference for the distorted image, which compensates the absence of true reference information and simulates the behavior of the human visual system. Furthermore, due to the design of domain awareness, it determines the distortion type of the input distorted image.

1) Latent Feature Loss: The latent feature loss aims to penalize the latent representations from two aspects. To learn disentangled representation across domains, we use the first term $\mathcal{L}_{kl} = KL(q(z|x)\|p(z))$ to calculate the Kullback-Leibler divergence, which makes the latent code z close to a prior Gaussian distribution $p(z)$. However, this term alone cannot guarantee the disentanglement of domain-specific information from the latent space, since the generator recovers the distorted-free images simply from the representation z without using any domain information.

To address the above problem and achieve simultaneous training of multiple domains with a single model, we extend the loss by adding a domain classification loss term, to eliminate the domain-specific information from the representation z. We assign a unique label for each domain, and introduce an auxiliary domain classifier which tries to distinguish the latent representations z from different domains.

Typically, latent representation z can be learned directly from the images through an encoder-decoder architecture. The encoder extracts features from images and the decoder uses these features to reconstruct the images. The latent feature loss is used to force the encoder to disentangle the content and the degradation information. Without the latent feature loss, the model can still learn a latent representation from the images, but the obtained representation would be entangled as shown in the left-hand panels of Figure 2 and Figure 6, since no constraints are added on the disentanglement process.

Different from [31], which aims to translate facial attributes among domains, image restoration transfers *several* domains of different distortions into one *single* distortion-free domain. Thus, we assign labels to representations instead of images. More precisely, we introduce a domain discriminator D_c, which takes the latent representation z in domain c as input and aims to distinguish the predicted domain code v_c from its real domain code. In contrast, the encoder E tries to confuse D_c from predicting the correct domain. The second term of latent feature loss, named domain classification loss, can be defined as

$$\mathcal{L}_E^{cls} = -\mathcal{L}_{D_c}^{cls} = -\mathbb{E}[-\log P(v_c|E(x_c)], \quad (1)$$
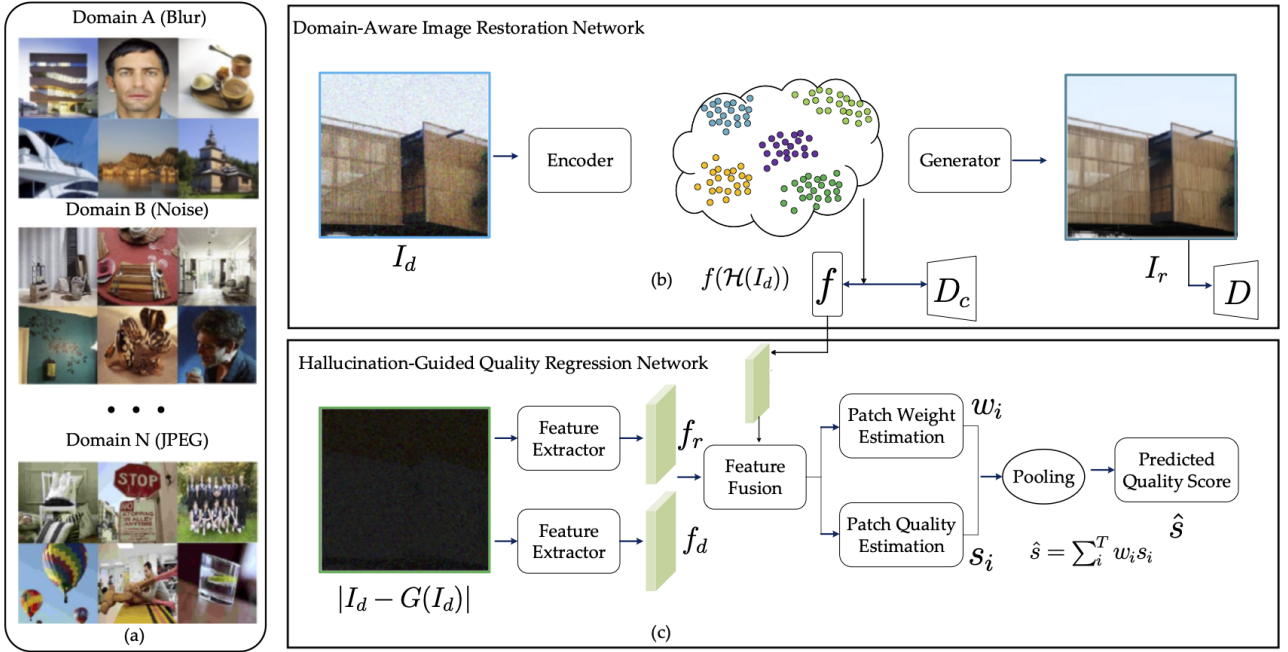
Fig. 3. **The framework of the proposed DA-IQA.** Given a dataset that contains several collections of images with different degradations (a), referred to as n domains, we first randomly select a distorted image I_d together with its specific domain label c from a certain domain $\mathcal{D}$, the generator G of Domain-Aware Image Restoration Network (b) aims to separate the distortion d from I_d to get the restoration I_r, while the discriminator D tries to distinguish if the input image is real or fake and the domain discriminator D_c recognizes the domain category (D and D_c are omitted from the framework for simplicity). The original distorted image I_d and its discrepancy map $|I_d - G(I_d)|$ are fed into the Hallucination-Guided Quality Regression Network (c). The two images are first through feature extractors to get their corresponding features f_r and f_d. The two feature extractors share the same weights. Then f_r and f_d are concatenated with high-level domain-aware distortion representation $f(\mathcal{H}(I_d))$, which is extracted from the restoration network. The fused feature is regressed to a patch-wise quality and weight estimation. The score of an image is the weighted average of all scores of its patches.

where $z = E(x_c)$ with x_c denoting the input image, and P is the probability distribution over domains c, which is produced by the domain discriminator D_c. Ideally, the latent representation z should contain distortion-free image content information together with disentangled domain-specific degradation information.

2) Perceptual Loss: From our preliminary experiments, we observe some unpleasing artifacts in the restored images. Motivated by [13], we add a perceptual loss [33] $\mathcal{L}_P$ between the ground-truth images I_{gt} and the corresponding restored ones I_r:

$$\mathcal{L}_P = \mathbb{E}[\sum_i \frac{1}{N_i} \|\phi_i(I_r) - \phi_i(I_{gt})\|_1], \quad (2)$$

where ϕ_i denotes the ith layer of the feature maps extracted from the pre-trained VGG-19 network, N_i is the number of the selected layers. The features extracted from pre-trained deep networks contain rich semantic information and are widely used in learning-based image generation tasks. The perceptual loss encourages the decoder G to generate images that match ground-truth images perceptually.

Different from the per-pixel loss functions, the perceptual loss compares high level differences such as content and style discrepancies between images, instead of understanding differences at a pixel level. The perceptual loss is a commonly used loss function to provide accurate and photo-realistic results.

3) Adversarial Loss: The VAE architecture tends to generate blurry samples [34], which would not be desirable for practical use. To get satisfactory image restoration from latent representation, we additionally introduce an image discriminator D, which also enhances the ability of representation disentanglement from the latent space. We define the objective functions $\mathcal{L}_D^{adv}$ and $\mathcal{L}_G^{adv}$ for adversarial learning between image discriminator D and image generator G as

$$\begin{aligned} \mathcal{L}_D^{adv} &= \mathbb{E}[\log(D(\hat{x}))] + \mathbb{E}[\log(1 - D(x_c))], \\ \mathcal{L}_G^{adv} &= -\mathbb{E}[\log(D(\hat{x}))], \end{aligned} \quad (3)$$

where x and $\hat{x}$ denote the input image and its restoration, respectively.

4) Overall Training Loss: For stable training, high image quality and considerable diversity, we use the least-squares GAN [35] in our experiment. The total training loss functions of the encoder E, decoder G, image discriminator D and domain discriminator D_c are defined as

$$\begin{aligned} \mathcal{L}_E &= \lambda_1 \mathcal{L}_{kl} + \lambda_2 \mathcal{L}_P + \mathcal{L}_E^{cls}, & \mathcal{L}_{D_c} &= \mathcal{L}_{D_c}^{cls}, \\ \mathcal{L}_G &= \lambda_1 \mathcal{L}_{kl} + \lambda_2 \mathcal{L}_P + \mathcal{L}_G^{adv}, & \mathcal{L}_D &= \mathcal{L}_D^{adv}, \end{aligned} \quad (4)$$

where λ_1 and λ_2 are regularization parameters controlling the importance of losses.

D. Hallucination-Guided Quality Regression Network

As shown in Figure 3, the DA-Restore module generates the residual image for restoration. This residual image is different

| Category	Method	TID2013		LIVE	
		SROCC	PLCC	SROCC	PLCC
FR-IQA	PSNR	0.889	0.847	0.880	0.805
	SSIM [36]	0.856	0.867	0.918	0.780
	FSIM [37]	**0.963**	0.932	**0.952**	0.822
	VSI [38]	0.947	**0.939**	0.936	**0.853**
NR-IQA	DIVINE [1]	0.855	0.851	0.885	0.853
	BLIINDS-II [39]	0.877	0.841	0.931	0.930
	BRISQUE [2]	0.922	0.917	0.940	0.911
	CORNIA [7]	0.903	0.917	0.913	0.888
	CNN [5]	0.903	0.917	0.913	0.888
	CNN++ [16]	0.843	0.804	0.928	0.897
	DIQaM-NR [6]	0.933	0.909	0.960	0.972
	RAN [14]	0.948	**0.937**	0.972	0.968
	DA-IQA	**0.952**	0.929	**0.977**	**0.975**
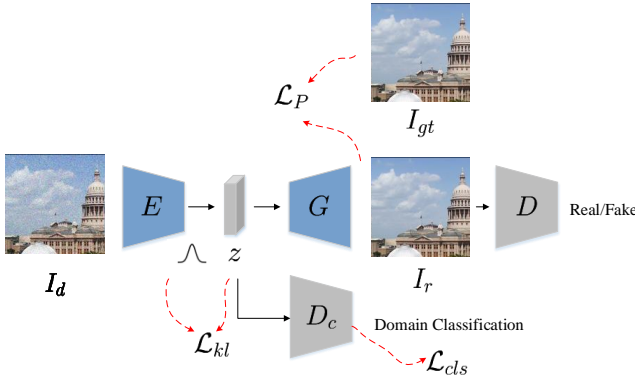
Fig. 4. **Overview of our proposed DA-Restore network.** This module takes the distorted image as input and aims to produce its corresponding restoration. Meanwhile, the model tries to disentangle the discriminative representations for different degradations.

from the concept of *error map* in FR-IQA, which represents pixel-wise error between the distorted image and the reference. To emphasize the difference, we refer to the residual image as *discrepancy map* and use the original distorted image together with the discrepancy map as input to acquire the desired quality score. The NR-IQA problem is now formulated as a regression by solving

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathcal{R}(I_d^i, |I_d^i - G(I_d^i)|), s_{gt}^i), \quad (5)$$

where $\mathcal{R}(\cdot)$ denotes the regression network for predicting the quality score, $|I_d^i - G(I_d^i)|$ is the discrepancy map generated by the DA-Restore module G, and s_{gt} represents the ground truth score.

Previous studies have shown that the quality results obtained by methods based on HVS are greatly affected by the eligibility of the restoration [13, 14]. An unqualified hallucination as the reference would introduce a large bias by deteriorating the gap between the distorted image and the restored one. To alleviate this drawback and stabilize the quality regression, we fuse the discrepancy information together with the high-level

information from the generative network as with [13]. Different from their work, ours uses the domain-specific distortion information.

Assume G has been trained. As shown in Figure 3, features of the original distorted image and its *discrepancy map*, f_r and f_d, are extracted by a CNN and are concatenated with the high-level domain-aware distortion representation $f(\mathcal{H}(I_d))$, which is extracted from the restoration network. $\mathcal{H}(I_d)$ denotes the feature maps of the distorted image, $f(\cdot)$ is a linear projection to ensure that the dimensions of $\mathcal{H}$ and $\mathcal{K}$ are equal, and $\mathcal{K}$ represents the feature maps that concatenate with $\mathcal{H}$. Then the fused feature is regressed to a patch-wise estimation of quality and weight estimation, denoted by s_i and w_i, respectively, and w_i is treated as the relative importance for each patch i.

The loss function of the patch quality estimation would be formulated as

$$\mathcal{L}_R = \frac{1}{T} \sum_{i=1}^{T} \|s_i - s_{gt}\|_1, \quad (6)$$

where $s_i = \mathcal{R}_2(f(\mathcal{H}(I_d)) \oplus \mathcal{K}(I_d, |I_d^i - G(I_d^i)|))$, $\mathcal{R}_2$ is the fully connected layers of $\mathcal{R}(\cdot)$, $\oplus$ denotes the concatenation operation, and T is the number of patches.

We assign different weights for the respective patches and use the normalized weights to estimate the quality score of the whole image. The reason of integrating a branch of weight estimation is that simply averaging all local quality scores does not consider the effect of spatial variance of relative image quality and perceptual relevance of local quality. The two branches are parallel and share the same dimension. The weight w_i of each input patch i is calculated by activating the output of weight estimation branch w' through a ReLU and adding a small stability term:

$$w_i = \max(0, w') + \epsilon. \quad (7)$$

The final image quality score $\hat{s}$ is thus

$$\hat{s} = \sum_{i}^{T} w_i s_i. \quad (8)$$

E. Network Architecture

For the domain-aware image restoration network, inspired by recent image translation studies, our generators follow an encoder-decoder architecture similar to that in [24]. Given depth $d = 5$, the ith layer of Encoder E operates on 4×4 spatial regions with a stride 2 and produces a feature map with size of $\{64 \times 2^{i-1}\}_{i=1}^{d}$, *i.e.*,16, 128, 256, 512, 1024, respectively. Each convolutional layer is followed by Instance Normalization [40], and Leaky ReLU are utilized. Generator G follows a reversed symmetrical architecture of E. Residual learning is adopted because it has been shown effective for image processing tasks and helpful on convergence. That is, the generator only learns the difference between the input image and the ground truth image. The domain discriminator D_c (discriminator with an auxiliary domain classifier) is built on top of the discriminator D.

For the hallucination-guided quality regression network, the domain-aware distortion representation is extracted from the well-trained restoration network and acts as compensation for IQA. The hallucination-guided quality regression network takes the distorted patch and the corresponding restored residual patch as input. Feature representations f_r and f_d are extracted by the same layers and fused with obtained feature map z. The fused feature vector is then fed into two branches for estimating perceptual score s_i and weight w_i, respectively.

IV. EXPERIMENTS

In this section, we conduct several experiments to test the performance of our proposed method on various IQA benchmarks. We pre-train DA-IQA on Waterloo Exploration [41] and perform cross validation on TID2013 [42] and LIVE [43].

A. Datasets and Evaluation Protocols

1) TID2013: TID2013 includes 25 distortion-free reference images and 3000 distorted images. These images are created from references with 24 types and 5 levels of distortions, ranging from additive Gaussian noise to Chromatic aberrations. Every image is annotated with Mean Opinion Scores (MOS), which is produced by several observers in subjective tests. The obtained MOS has to vary from 0 to 9 and its larger values correspond to better perceptual quality. Its wide range makes it one of the most comprehensive IQA databases.

2) LIVE: LIVE consists of 29 reference images and 779 distorted samples with 5 distortion types including Fast Fading, Gaussian Blur, White Noise, JPEG Compression and JP2K Compression. Each image is provided with Differential Mean Opinion Scores (DMOS), ranging from 0 to 100. Lower DMOS means higher perceptual quality. DMOS value of zero indicates the image is distortion-free.

3) Waterloo Exploration: Waterloo Exploration contains 4744 pristine natural images and 94880 distorted images. The distorted images are generated by MATLAB with four distortion types and five levels. Compared to TID2013 and LIVE, Waterloo Exploration has much larger amounts of images, thus it also has a great diversity of image content. The four types, *i.e.*, JPEG Compression, JP2K Compression, Gaussian Blur and White Noise, are also considered the most

common distortion types and are covered both in TID2013 and LIVE.

4) Evaluation Metrics: Following most previous works [5, 6, 14], the performances on above datasets are evaluated by two common metrics for model evaluation: the Linear Correlation Coefficient (LCC) and the Spearman's Rank Order Correlation Coefficient (SROCC). LCC is a measure of the linear correlation between the ground-truth and model prediction, which is defined as

$$\text{LCC} = \frac{\sum_{i=1}^{N} (y_i - \overline{y}_i)(\hat{y}_i - \overline{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{N} (y_i - \overline{y}_i)^2} \sqrt{\sum_{i=1}^{N} (\hat{y}_i - \overline{\hat{y}}_i)^2}}, \quad (9)$$

where $\overline{y}_i$ and $\overline{\hat{y}}_i$ denote the means of the ground truth and predicted score, respectively. SROCC measures the prediction monotonicity between the ground-truth and model prediction, which could be formulated as

$$\text{SROCC} = 1 - \frac{6 \sum_{i=1}^{N} (v_i - p_i)^2}{N(N^2 - 1)}, \quad (10)$$

where N represents the number of distorted images, v_i, p_i are the positions of $\hat{y}_i, \overline{\hat{y}}_i$ in the ranking sequences. The SROCC measures the monotonic relationship between the ground-truth and estimated IQAs while the PLCC is a measure of the linear correlation between the ground-truth and predicted quality scores. Higher SROCC score means higher monotonicity and higher PLCC score represents higher linear correlation, between the ground-truth and predicted quality scores.

B. Implementation Details

All the training images are randomly sampled from the original images of size 256×256 with stride 96. We train our model with Pytorch on the GeForce GTX 1080Ti with a batch size of 32. We apply Adam [44] solver with parameters of learning rate 0.0002, $\beta_1 = 0.5, \beta_2 = 0.999$. For domain-aware image restoration, the weights λ_1, λ_2 are set as 5 and 100, respectively. For hallucination-guided quality regression, we split the TID2013 and LIVE datasets into 6:2:2 for training, validation and test, respectively. Considering that perceptual scores are not available in Waterloo Exploration, we provide a score and weight for every distorted image patch by performing FSIM [37], which is one of the state-of-the-art FR-IQA metrics. These scores and weights serve as the ground truth label in the following training process as in [14]. During the testing of image quality assessment, we extract overlapped patches from each test image at a fixed stride and calculate the weighted scores of all patches as the final quality score.

C. Cross Validation on TID2013 and LIVE

We firstly train the domain-aware image restoration module on Waterloo Exploration, then train the hallucination-guided quality regression on TID2013 and LIVE to perform cross validation, respectively. Since Waterloo Exploration contains only four distortion types, we train and test DA-IQA on these following types: Gaussian Blur, White Noise, JPEG and JP2K, following the same setting as in [14]. As shown in Table I, the proposed DA-IQA performs better than almost all the
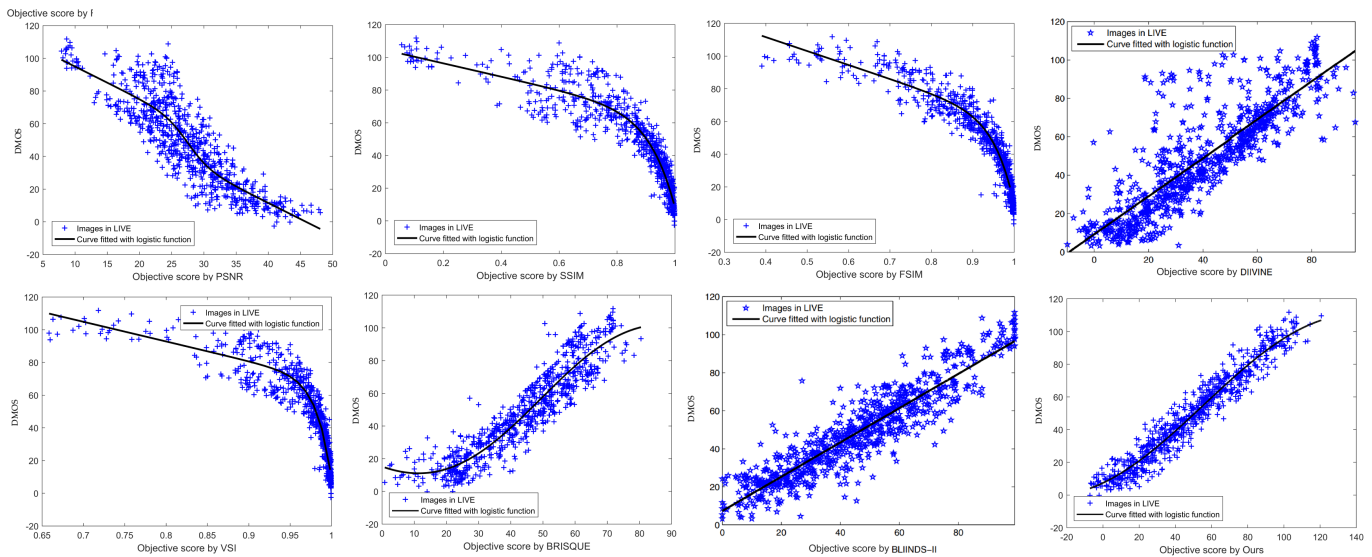
Fig. 5. The scatter plot of MOS values with respect to objective values on the LIVE dataset. The blue "+" represents one distorted image and the black curves are obtained in the curve fitting process. The blue "+" of our method gather evenly and close to the black curve and the curve also exhibits almost a straight line, which manifests the better correlation between the scores given by our method and the subjective judgements for the image quality.

other state-of-the-art FR-IQA and NR-IQA methods in terms of both PLCC and SROCC. Furthermore, our model outputs simultaneously quality estimation and distortion identification. As shown in Table II, our method also wins the distortion identification task compared with other distortion-identification IQA methods [4, 16]. Table III shows the confusion matrices produced by our method on the LIVE dataset. The column and the raw contain ground truths and predicted distortion types, respectively.

For visualization, we also provide the scatter plot of subjective MOS values with respect to objective values on the LIVE database in Figure 5, in which we denote the distorted images with blue "+" and the black curves are obtained in the curve fitting process as in [45]. One can see the blue "+" of our method gather evenly and close to the black curve and the curve also exhibits almost a straight line, which manifests the better correlation between the scores given by our method and the subjective judgements for the image quality.

TABLE II
PERFORMANCE OF DISTORTION IDENTIFICATION ON 5 COMMON DISTORTIONS OF LIVE AND 24 TYPES OF DISTORTION ON TID2013.

Method	Accuracy	
	LIVE	TID2013
CNN++ [16]	0.925	0.819
MEON [4]	0.912	0.859
DA-IQA	**0.942**	**0.937**

D. Results on More Distortion Types

We show results on four distortions in Section IV-C, to demonstrate the scalability of our domain-aware framework when handling more distortion types, we further train and test on the full TID2013 dataset. For pre-training, as in the

TABLE III
THE CONFUSION MATRICES PRODUCED BY OUR METHOD ON THE LIVE DATASET. THE COLUMN AND THE RAW CONTAIN GROUND TRUTHS AND PREDICTED DISTORTION TYPES, RESPECTIVELY.

	Category	JP2K	JPEG	WN	BLUR	Pristine
	JP2K	**0.915**	0.010	0.000	0.023	0.032
	JPEG	0.048	**0.919**	0.000	0.022	0.011
LIVE	WN	0.000	0.000	**1.000**	0.000	0.000
	BLUR	0.059	0.007	0.000	**0.926**	0.008
	Pristine	0.007	0.013	0.000	0.034	**0.950**

previous studies [4, 46], we reproduce 17 out of a total of 24 distortion types in TID2013 and apply them to the 840 high-quality images. For the distortions we did not reproduce (*i.e.*, #3, #4, #12, #13, #20, #21, #24), we fine-tune from other distortions. We follow the experimental setting as used in [46] and the average SROCC of all the experiments performed 10 times is reported in Table IV, where *ALL* means testing all distortions together. From Table IV, several patterns can be observed. First, our proposed DA-IQA outperforms previous models by a clear margin, which strongly demonstrates the effectiveness of domain fingerprints. Second, the DA-Restore has learned domain-aware features, which are fed into the subsequent module to get the desired score. To some extent, image restoration and image quality assessment are inherently consistent. The features learned during the restoration can not only be used to reconstruct distort-free images but also contain the degradation type and level information of given images that can be further processed to get the quality score. However, it is hard to obtain satisfactory results by directly training a deep network on IQA data due to the data scarcity. Thus we pre-trained the DA-Restore module on other large-scale high-quality datasets to learn the domain-aware features of different degradations (domain fingerprints) and fine-tuned the quality

regression module on accurate IQA scores.

E. Feature Disentanglement Visualization

To demonstrate the ability for disentanglement and transfer-ability of learned features, many works [24, 47] conduct feature visualization with t-SNE [32]. Similarly, we simultaneously perform feature disentanglement of distortions at different domains and show the results in Figure 6. To be more specific, we trained the DA-Restore module alone with the DIV2K dataset [48] aiming to restore images of multiple degradations in one model. The DIV2K [48] is often used in the image restoration task as the augmented training dataset. Each color indicates a different domain, *i.e.*, noisy images with σ 15, 25, 30, 50 and 70, JPEG compressed images with quality factor 10 and 20, and low-resolution images with factor 2, 3 and 4. As shown, features of images from different domains are discriminated significantly well.
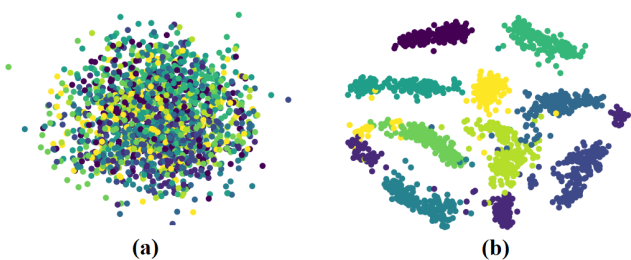

(a)   **(b)**

Fig. 6. Feature visualization (a) without and (b) with domain-aware mechanism. Better view in color.

F. Fingerprint Visualization

In Section II-B and Section IV-E, we describe how to learn and analyze the domain fingerprint which are **implicitly** represented in the feature domain. To help better understand the domain fingerprints, following the similar spirit to [27, 28], we **explicitly** represent them in the image domain by introducing the image fingerprint F_{im}^{I} and the model fingerprint F_{mod}^{d}. The image fingerprints are defined as the reconstruction residual, and the domain fingerprints are learned simultaneously from each domain. The response is simply defined as pixel-wise multiplication of two normalized images $F_{im}^{I} \odot F_{mod}^{d}$.

As shown in Figure 7, we can see that images from different degradations possess different fingerprint patterns. Their pair-wise interactions are shown as the confusion matrix. Different from the confusion matrix in Table III, which compares final predicted distortion types and ground truth labels, this confusion matrix is image fingerprint responding to model fingerprint. Ideally, the image fingerprints should maximize responses only to their own domain fingerprints. The results in Figure 7 are not visually and perceptually obvious. This can be due to two reasons: a) Incomplete feature disentanglement. As emphasized in [49], during feature disentanglement, the degradation information extracted from a single image contains some more information rather than only generalized fingerprint of image collections. In the training phase, the model may extracts some content features as degradation fingerprints

incorrectly, which means that the fingerprints still contain some general shared image content. b) The reconstruction error. We define the reconstruction residual as the image fingerprint. It is convenient for visualizing and understanding implicit features. However, the reconstruction results of our DA-Restore module could not treat to the same extent as the final restoration. Our goal is to learn and analyze useful feature for image quality assessment instead of restore images. Nonetheless, an improved strategy for thorough disentanglement [49] and reconstruction is our future work.
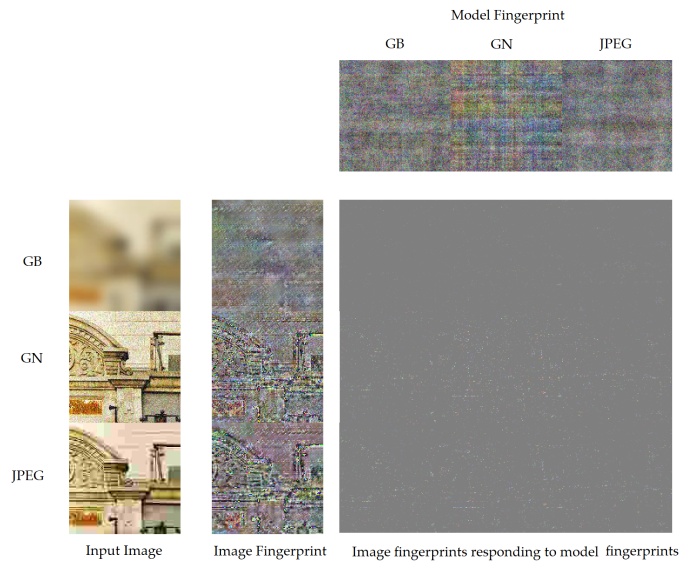

Fig. 7. Visualization of model and image fingerprint samples. Their pairwise interactions are shown as the confusion matrix. It is simply defined as pixel-wise multiplication of two normalized images.

G. Ablation Study

To demonstrate the efficacy of the key components of our method for the performance, we conduct several ablation experiments on TID2013, in which we remove perceptual loss, adversarial learning, high-level semantic fusion, or domain classification and test the performance of the remaining framework by comparing both SROCC and LCC results, as shown in Table V.

1) Domain Classification: To show how "domain classifica-tion" contributes to the performance, we remove the domain classification mechanism. Domain classification mechanism is a crucial component to make our approach "domain-aware". As illustrated in Figure 6, without domain classification, the representation disentanglement would be only separation of content and degradation. The representations of different degradations are still entangled and no discriminative features of certain degradation are learned. The results in Table V show that the discrimination of different degradations can actually boost the performance of the IQA task.

2) High-Level Semantic Fusion: The two aforementioned mechanisms contribute to the IQA performance by indirectly improving the quality of restored reference. When the restora-tion is unqualified, a large bias would be introduced and lead

TABLE IV
PERFORMANCE EVALUATION (SROCC) ON THE ENTIRE TID2013 DATASET.

Method	#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13
BLIINDS-II [39]	0.714	0.728	0.825	0.358	0.852	0.664	0.780	0.852	0.754	0.808	0.862	0.251	0.755
BRISQUE [2]	0.630	0.424	0.727	0.321	0.775	0.669	0.592	0.845	0.553	0.742	0.799	0.301	0.672
CORNIA [7]	0.341	-0.196	0.689	0.184	0.607	-0.014	0.673	0.896	0.787	0.875	0.911	0.310	0.625
RankIQA	0.667	0.620	0.821	0.365	0.760	0.736	0.783	0.809	0.767	0.866	0.878	0.704	0.810
DA-IQA	0.903	0.801	0.905	0.714	0.891	0.879	0.921	0.912	0.897	0.919	0.925	0.609	0.651
Method	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	ALL	
BLIINDS-II [39]	0.081	0.371	0.159	-0.082	0.109	0.699	0.222	0.451	0.815	0.568	0.856	0.550	
BRISQUE [2]	0.175	0.184	0.155	0.125	0.032	0.560	0.282	0.680	0.804	0.715	0.800	0.562	
CORNIA [7]	0.161	0.096	0.008	0.423	-0.055	0.259	0.606	0.555	0.592	0.759	0.903	0.651	
RankIQA	0.597	0.622	0.268	0.613	0.662	0.619	0.644	0.800	0.779	0.629	0.859	0.780	
DA-IQA	0.477	0.695	0.438	0.674	0.709	0.852	0.713	0.897	0.808	0.774	0.868	0.828	

TABLE V
ABLATION EXPERIMENT ON TID2013. "W/O M" MEANS OUR MODEL
WITHOUT COMPONENT M.

Ablation	TID2013	
	PLCC	SROCC
Proposed (Ours)	**0.929**	**0.952**
w/o Domain Classification	0.913	0.934
w/o Semantic Fusion	0.917	0.902
w/o Perceptual Loss	0.883	0.876
w/o Adversarial Learning	0.864	0.859

to deterioration of the gap between the distorted image and the restored one. Thus we design high-level semantic fusion to alleviate this drawback and stabilize the quality regression. To show its impact, we remove the fusion module, and the ablation results shown in Table V demonstrate the validity.

3) Perceptual Loss: We first evaluate the effect of perceptual loss. The ablated model obtains an obvious performance decline by removing the perceptual loss since such a loss helps to restoration at the training process.

4) Adversarial Learning: To explore how adversarial learning contributes to the restoration performance, we further evaluate the model without image discriminator and adversarial loss. Removal of adversarial learning leads to significant performance decline since the discriminator no longer propelled the generator.

V. CONCLUSION

In this paper, we propose the domain-aware no-reference image quality assessment (DA-IQA). The proposed DA-IQA leverages domain fingerprints for image quality assessment. These domain fingerprints reveal the disentangled discriminative feature representations of different degradations. Benefiting from the design of domain-aware network architecture, our method is also able to identify the distortion type of an image, and use the determined distortion type and quality score to characterize the image quality. Experiments on various standard IQA datasets have shown its superiority over state-of-the-art IQA methods.

REFERENCES

[1] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, 2011.

[2] 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.

[3] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, 2015.

[4] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, 2018.

[5] L. Kang, P. Ye, Y. Li, and D. S. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1733–1740.

[6] S. Bosse, D. Maniry, K. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.

[7] 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.

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

[9] P. Zhang, W. Zhou, L. Wu, and H. Li, "SOM: Semantic obviousness metric for image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2394–2402.

[10] 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.

[11] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, 2012.

[12] 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.

[13] K. Lin and G. Wang, "Hallucinated-IQA: No-reference image quality assessment via adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 732–741.

[14] H. Ren, D. Chen, and Y. Wang, "RAN4IQA: Restorative adversarial nets for no-reference image quality assessment," in *Proc. Conf. Arti. Intell.*, 2018, pp. 7308–7314.

[15] H. Lim, H. G. Kim, and Y. M. Ra, "VR IQA NET: Deep virtual reality image quality assessment using adversarial learning," in *Proc. Conf. Acoustics, Speech and Signal Process.*, 2018, pp. 6737–6741.

[16] L. Kang, P. Ye, Y. Li, and D. S. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," in *Proc. Conf. Image Process.*, 2015, pp. 2791–2795.

[17] C. Huang, T. Jiang, and M. Jiang, "Encoding distortions for multi-task full-reference image quality assessment," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2019, pp. 1864–1869.

[18] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology-Paris*, vol. 100, pp. 70–87, 2006.

[19] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, 2012.

[20] E. L. Denton and V. Birodkar, "Unsupervised learning of disentangled representations from video," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4414–4423.

[21] T. Xiao, J. Hong, and J. Ma, "ELEGANT: Exchanging latent encodings with GAN for transferring multiple face attributes," in *Proc. Eur. Conf. Comput. Vis.*, September 2018, pp. 172–187.

[22] H. Lee, H. Tseng, J. Huang, M. Singh, and M. Yang, "Diverse image-to-image translation via disentangled representations," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 36–52.

[23] X. Huang, M. Liu, S. J. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 179–196.

[24] A. H. Liu, Y. Liu, Y. Yeh, and Y. F. Wang, "A unified feature disentangler for multi-domain image translation and manipulation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2595–2604.

[25] B. Lu, J.-C. Chen, and R. Chellappa, "Unsupervised domain-specific deblurring via disentangled representations," *arXiv preprint arXiv:1903.01594*, 2019.

[26] M. Chen, J. J. Fridrich, M. Goljan, and J. Lukás, "Determining image origin and integrity using sensor noise," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 1, pp. 74–90, 2008.

[27] F. Marra, D. Gragnaniello, L. Verdoliva, and G. Poggi, "Do GANs leave artificial fingerprints?" in *IEEE Conf. Multimedia Information Processing and Retrieval*, 2019, pp. 506–511.

[28] N. Yu, L. S. Davis, and M. Fritz, "Attributing fake images to GANs: Learning and analyzing GAN fingerprints," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019.

[29] S. Choudhary, P. Srivastava, L. H. Ungar, and J. Sedoc, "Domain aware neural dialog system," *arXiv preprint arXiv:1708.00897*, 2017.

[30] A. Saha, M. M. Khapra, and K. Sankaranarayanan, "Towards building large scale multimodal domain-aware conversation systems," in *Proc. Conf. Arti. Intell.*, 2018, pp. 696–704.

[31] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8789–8797.

[32] P. E. Rauber, A. X. Falcão, and A. C. Telea, "Visualizing time-dependent data using dynamic t-SNE," in *Euro. Conf. Vis.*, 2016, pp. 73–77.

[33] J. Johnson, A. Alahi, and F. F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[34] S. Zhao, J. Song, and S. Ermon, "Towards deeper understanding of variational autoencoding models," *arXiv preprint arXiv:1702.08658*, 2017.

[35] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2794–2802.

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

[37] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.

[38] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, 2014.

[39] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, 2012.

[40] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[41] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004–1016, 2017.

[42] N. N. Ponomarenko, L. Jin, O. Ieremeiev, V. V. Lukin, K. O. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C. J. Kuo, "Image database TID2013: peculiarities, results and perspectives," *Sig. Proc.: Image Comm.*, vol. 30, pp. 57–77, 2015.

[43] 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.

[44] D. Kingma and J. Ba, "Adam: A method for stochastic

optimization," *Computer Science*, 2014.

[45] Y. Liu, K. Gu, S. Wang, D. Zhao, and W. Gao, "Blind quality assessment of camera images based on low-level and high-level statistical features," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 135–146, 2019.

[46] X. Liu, J. van de Weijer, and A. D. Bagdanov, "RankIQA: Learning from rankings for no-reference image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1040–1049.

[47] J. Hoffman, S. Guadarrama, E. Tzeng, R. Hu, J. Donahue, R. B. Girshick, T. Darrell, and K. Saenko, "LSDA: large scale detection through adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3536–3544.

[48] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, July 2017.

[49] W. Xia, Y. Yang, and J.-H. Xue, "Unsupervised multi-domain multimodal image-to-image translation with explicit domain-constrained disentanglement," *arXiv preprint arXiv:abs/1911.00622*, 2019.

Weihao Xia received the M.Eng. degree in Control Engineering from Tsinghua University in 2019 and the B.Eng. degree from the School of Information Sciences and Technology, Sun Yat-sen University in 2016. His research interests include computer vision and computer graphics.

Yujiu Yang received the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences. He is an associate professor in Tsinghua Shenzhen International Graduate School, Tsinghua University. His research interests include natural language processing and computer vision.

Jing-Hao Xue received the Dr.Eng. degree in signal and information processing from Tsinghua University in 1998 and the Ph.D. degree in statistics from the University of Glasgow in 2008. He is an associate professor in the Department of Statistical Science at University College London. His research interests include statistical classification, high-dimensional data analysis, computer vision and pattern recognition.

Jing Xiao is the Group Chief Scientist of Ping An Insurance (Group) Company of China, LTD. He received his Ph.D. degree from the School of Computer Science, Carnegie Mellon University, and has published over 80 academic papers and 100 granted US patents. Before joining Ping An, he worked as Principal Applied Scientist Lead in Microsoft Corp. and Manager of Algorithm Group in Epson Research and Development, Inc. Dr. Xiao started R&D in artificial intelligence and related fields since 1995, covering a broad range of application areas such as healthcare, autonomous driving, 3D printing and display, biometrics, web search, and finance. He is now leading research and development in AI-related technologies and their applications on finance, healthcare, and smart-city in Ping An.